ACKNOWLEDGMENTS

In addition to the Editorial Board of Consulting Editors, the following individuals ted in the review of manuscripts submitted to the *Journal* during calendar year 1970. Editor gratefully acknowledges their contributions.

Preston S. Abbott
Malcolm D. Arnoult
William J. Baker
C. J. Bartlett
David P. Campbell
John P. Campbell
Ronald P. Carver
Royer F. Cook
John A. Creager
Terry Eisenberg
Alfred J. Farina, Jr.
James L. Farr
Fred E. Fiedler
Robert H. Fosen
Robert M. Guion
J. Richard Hackman
Milton D. Hakel
John L. Holland
Herbert Horowitz
Donald P. Horst
Charles L. Hulin
Irving L. Janis
William A. Johnston
Abraham K. Korman
John T. Lanzetta
Edward E. Lawler
Eugene C. Mayfield
Angelo Mirabella
Thomas I. Myers
Brian S. O'Leary
Seymour Rosenberg
Nancy F. Russo
Paul A. Schwarz
David I. Sheppard
Patricia C. Smith
Paul Spector
Matilde K. Stephenson
Robert W. Stephenson
Herbert Stone
Warren H. Teichner
Frederick C. Thorne
Victor H. Vroom

LANCASTER PRESS, INC., LANCASTER, PA.

Contents of Volume 55

Abrahams, Norman M., Alf, Edward F., & Wolfe, John H. **Taylor-Russell Tables for Dichotomous Criterion Variables** 449
Abrahams, Norman M. See Lau, Alan W.
Abrams, Charles. See Taub, Harvey A.
Alderfer, Clayton P. **Effect of Individual, Group, and Intergroup Relations on Attitudes toward a Management Development Program** 302
Aleamoni, Lawrence M. See Kawash, Mary B.
Alexander, Judith F. See Laughery, Kenneth R.
Alf, Edward F. See Abrahams, Norman M.
Allen, John C. See Goldstein, Irwin L.
Alvares, Kenneth M. See Graen, George.
Alvares, Kenneth M. See Orris, James Burdeane.
Andrews, I. R. See Valenzi, E. R.
Armstrong, Thomas B. **Job Content and Context Factors Related to Satisfaction for Different Occupational Levels** 57
Ash, Phillip. **Screening Employment Applicants for Attitudes toward Theft** 161
Athelstan, Gary T., & Paul, Gerald J. **New Approach to the Prediction of Medical Specialization: Student-Based Strong Vocational Interest Blank Scales** 80
Bailey, Larry J. See Baldwin, Thomas S.
Baldwin, Thomas S., & Bailey, Larry J. **Readability of Technical Training Materials Presented on Microfiche versus Offset Copy** 37
Barrett, Gerald V. See Giles, Brian A.
Bishop, Ronald C., & Hill, James W. **Effects of Job Enlargement and Job Change on Contiguous but Nonmanipulated Jobs as a Function of Workers' Status** 175
Blakeney, Roger N., & MacNaughton, John F. **Effects of Temporal Placement of Unfavorable Information on Decision Making during the Selection Interview** 138
Blood Milton R. See Fogli, Lawrence.
Blood, Milton R. **The Validity of Importance** 487
Brainin, Uri. See Gannon, Martin J.
Brennan, Thomas F. See Gluskinos, Ury.
Bruvold, William H. **Affective Response toward Users of Reclaimed Water** 28
Campbell, David P. See Johansson, Charles B.
Campbell, David P. See Rossmann, Jack E.
Campbell, Douglas B. **Relative Influence of Job and Supervision on Shared Worker Attitudes** 521
Carlson, Robert E. **Effect of Interview Information in Altering Valid Impressions** 66
Carroll, Stephen J., Jr. See Gannon, Martin J.
Carver, Ronald P. **Pupil Dilation and Its Relationship to Information Processing during Reading and Listening** 126
Castle, M. See Eysenck, H. J.
Chaison, Gary N. **Changes in Authorship, Affiliation, and Research Support Patterns of *Journal of Applied Psychology* Articles, 1917–1969** 484
Cherrington, David J., Reitz, H. Joseph, & Scott, William E., Jr. **Effects of Contingent and Noncontingent Reward on the Relationship between Satisfaction and Task Performance** 531
Clark, Barbara. See Schuster, Jay R.
Collins, William E., Schroeder, David J., Gilson, Richard D., & Guedry, Frederick E., Jr. **Effects of Alcohol Ingestion on Tracking Performance during Angular Acceleration** 559
Cook, Royer F., Fosen, Robert H., & Pacht, Asher. **Pornography and the Sex Offender: Patterns of Previous Exposure and Arousal Effects of Pornographic Stimuli** 503
Craft, John L. See Simon, J. Richard.
Cranny, C. J. See Zedeck, Sheldon.
Cummings, L. L. Schwab, Donald P., & Rosen, Marc. **Performance and Knowledge of Results as Determinants of Goal Setting** 526
Cummins, Robert C. **Relationship of Initiating Structure and Job Performance as Moderated by Consideration** 489
Donnelly, James H. See Ivancevich, John M.
Dorfman, Peter W., & Goldstein, Irwin L. **Spatial and Temporal Information as Cues in a Time-Sharing Task** 554
Dumas, Neil S., & Muthard, John E. **Job Analysis Method for Health-Related Professions: A Pilot Study of Physical Therapists** 458
Elton, Charles F. **Interaction of Environment and Personality: A Test of Holland's Theory** 114
Eysenck, H. J., & Castle, M. **Comparative Study of Artists and Nonartists on the Maitland Graves Design Judgment Test** 389
Fenelon, James R., & Megargee, Edwin I. **Influence of Race on the Manifestation of Leadership** 353
Fiedler, Fred E. **Note on the Methodology of the Graen, Orris, and Alvares Studies Testing the Contingency Model** 202
Fiedler, Fred E., Mitchell, Terrence, & Triandis, Harry C. **The Culture Assimilator: An Approach to Cross-Cultural Training** 95
Field, T. W., Simpkins, W. S., Browne, R. K., & Rich, P. **Identifying Patterns of Teacher Behavior from Student Evaluations** 466

Filley, Alan C. See House, Robert J.
Fine, B. D. **Comparison of Work Groups with Stable and Unstable Membership** 170
Fischl, M. A. See Siegel, Arthur I.
Fogli, Lawrence, Hulin, Charles L., & Blood, Milton, R. **Development of First-Level Behavioral Job Criteria** 3
Fosen, Robert H. See Cook, Royer F.
Fozard, James L., & Nuttall, Ronald L. **General Aptitude Test Battery Scores for Men Differing in Age and Socioeconomic Status** 372
Friedlander, Frank, & Greenberg, Stuart. **Effects of Job Attitudes, Training, and Organization Climate on Performance of the Hard-Core Unemployed** 287
Gannon, Martin J. **Sources of Referral and Employee Turnover** 226
Gannon, Martin J., & Brainin, Uri. **Factors Related to Acceptance and Refusal of Employment among Temporary Help Applicants** 493
Gannon, Martin J., Nothern, Joseph C., & Carroll, Stephen J., Jr. **Characteristics of Nonrespondents among Workers** 586
Giles, Brian A., & Barrett, Gerald V. **Utility of Merit Increases** 103
Gilson, Richard D. See Collins, William E.
Gluskinos, Ury, & Brennan, Thomas F. **Selection and Evaluation Procedure for Operating Room Personnel** 165
Goldstein, Irwin L. **The Application Blank: How Honest Are the Responses?** 491
Goldstein, Irwin L. See Dorfman, Peter W.
Goldstein, Irwin L., & Allen, John C. **Effects of Irrelevant Stimuli on the Processing of Information in Complex Displays** 110
Goldstein, Irwin L., & Mobley, William H. **Error and Variability in the Visual Processing of Dental Radiographs** 549
Goodale, James G. See Wollack, Stephen.
Graen, George, Orris, James Burdeane, & Alvares, Kenneth M. **Contingency Model of Leadership Effectiveness: Some Experimental Results** 196
Graen, George, Orris, James Burdeane, & Alvares, Kenneth M. **Contingency Model of Leadership Effectiveness: Some Methodological Issues** 205
Greenberg, Stuart. See Friedlander, Frank.
Grigaliunas, Benedict S., & Herzberg, Frederick. **Relevancy in the Test of Motivator-Hygiene Theory** 73
Guedry, Frederick E., Jr. See Collins, William E.
Gujarati, Damodar N. See House, Robert J.
Hackman, J. Richard. See Scheflen, Kenneth C.
Haddock, Rafael A. See Jacoby, Jacob.
Hakel, Milton D. **Similarity of Post-Interview Trait Rating Intercorrelations as a Contributor to Interrater Agreement in a Structured Employment Interview** 443
Herzberg, Frederick. See Grigaliunas, Benedict S.
Hill, James W. See Bishop, Ronald C.
House, Robert J., Filley, Alan C., & Gujarati, Damodar N. **Leadership Style, Hierarchical Influence, and the Satisfaction of Subordinate Role Expectations: A Test of Likert's Influence Proposition** 422
Hulin, Charles L. See Fogli, Lawrence.
Hulin, Charles L., & Waters, L. K. **Regression Analysis of Three Variations of the Two-Factor Theory of Job Satisfaction** 211
Hundal, P. S. **A Study of Entrepreneurial Motivation: Comparison of Fast- and Slow-Progressing Small-Scale Industrial Entrepreneurs in Punjab, India** 317
Ivancevich, John M., & Donnelly, James H. **Job Offer Acceptance Behavior and Reinforcement** .. 119
Jacoby, Jacob. **Multiple-Indicant Approach for Studying New Product Adopters** 384
Jacoby, Jacob, Olson, Jerry C., & Haddock, Rafael A. **Price, Brand Name, and Product Composition Characteristics as Determinants of Perceived Quality** 570
Jain, Subhash. See Misra, Sasi.
Johannson, Charles B., & Campbell, David P. **Stability of the Strong Vocational Interest Blank for Men** 34
Johnson, Richard W. **Comparability of Old and Revised Forms of the Strong Vocational Interest Blank for Women** 50
Kawash, Mary B., & Aleamoni, Lawrence M. **Effect of Personal Signature on the Initial Rate of Return of a Mailed Questionnaire** 589
Kennedy, John L. **The System Approach: A Preliminary Exploratory Study of the Relation between Team Composition and Financial Performance in Business Games** 42
Korman, Abraham K. **Environmental Ambiguity and Locus of Control as Interactive Influences on Satisfaction** 339
Korman, Abraham K. **Expectancies as Determinants of Performance** 218
Landy, Frank J. **Motivational Type and the Satisfaction-Performance Relationship** 406
Lane, Alan B. See Laughery, Kenneth R.
Lau, Alan W., & Abrahams, Norman M. **Stability of Vocational Interests within Nonprofessional Occupations** 143
Laughery, Kenneth R., Alexander, Judith F., & Lane, Alan B. **Recognition of Human Faces: Effects of Target Exposure Time, Target Position, Pose Position, and Type of Photograph** 477
Lawler, Edward E., III. See Scheflen, Kenneth C.

Leonard, Skipton, & Weitz, Joseph. Task Enjoyment and Task Perseverance in Relation to Task Success and Self-Esteem 414
Levine, Edward L., & Weitz, Joseph. Relationship between Task Difficulty and the Criterion: Should We Measure Early or Late? 512
Lewin, Arie Y., Dubno, Peter, & Akula, William G. Face-to-Face Interaction in the Peer-Domination Process 495
Lips, Orville J. See Rossmann, Jack E.
MacNaughton, John F. See Blakeney, Roger N.
Megargee, Edwin, I. See Fenelon, James R.
Mehler, Anne. See Wicker, Allan W.
Miskimins, R. W. See Searls, Donald J.
Misra, Sasi, & Jain, Subhash. Effect of Fittingness, Type of Goods, and Type of Slogan on Brand Awareness 580
Mitchell, Terrence. See Fiedler, Fred E.
Mobley, William H. See Goldstein, Irwin L.
Mulder, Franz. Characteristics of Violators of Formal Company Rules 500
Muthard, John E. See Dumas, Neil S.
Nothern, Joseph C. See Gannon, Martin J.
Nuttall, Ronald L. See Fozard, James L.
O'Dell, Jerry W. Method for Detecting Random Answers on Personality Questionnaires 380
Olson, Jerry C. See Jacoby, Jacob.
Orris, James Burdeane. See Graen, George.
Pacht, Asher. See Cook, Royer F.
Pahl, Bernd. See Vroom, Victor H.
Paul, Gerald J. See Athelstan, Gary T.
Dino, Christopher J. Relation of a Trainability Index to T-Group Outcomes 439
Putz, Vernon. See Smith, Karl U.
Putz, Vernon, & Smith, Karl U. Human Factors in Operating Systems Related to Delay and Displacement of Retinal Feedback 9
Raubenheimer, I. van W., & Tiffin, Joseph. Personnel Selection and the Prediction of Error 229
Reitz, H. Joseph. See Cherrington, David J.
Roach, Darrell E. Double Cross-Validation of a Weighted Application Blank over Time 157
Roach, Darrell. See Waters, L. K.
Rogers, Miles. See Schuster, Jay R.
Rosen, Hjalmar, & Turner, John. Effectiveness of Two Orientation Approaches in Hard-Core Unemployed Turnover and Absenteeism 296
Rosen, Marc. See Cummings, L. L.
Rosenbaum, Leonard L., & Rosenbaum, William B. Morale and Productivity Consequences of Group Leadership Style, Stress, and Type of Task 343
Rosenbaum, William B. See Rosenbaum, Leonard L.
Rossmann, Jack E., Lips, Orville J., & Campbell, David P. Vocational Interests of Political Scientists 135
Saleh, Shoukry D. Anxiety as a Function of Intrinsic-Extrinsic Job Orientation, the Presence or Absence of Observers, and Task Difficulty 543
Scheflen, Kenneth, C., Lawler, Edward E., III, & Hackman, J. Richard. Long-Term Impact of Employee Participation in the Development of Pay Incentive Plans: A Field Experiment Revisited 182
Schremser, Robert. See Smith, Karl U.
Schroeder, David J. See Collins, William E.
Schuster, Jay R., Clark, Barbara, & Rogers, Miles. Testing Portions of the Porter and Lawler Model regarding the Motivational Role of Pay 187
Schwab, Donald P. See Cummings, L. L.
Scott, William, E., Jr. See Cherrington, David J.
Searls, Donald J., Wilson, Lowell T., & Miskimins, R. W. Development of a Measure of Unemployability among restored Psychiatric Patients 223
Siegel, Arthur I., & Fischl, M. A. Dimensions of Visual Information Displays 470
Simon, J. Richard, & Craft, John L. Communicating Directional Information with an Auditory Display 241
Slocum, John W., Jr. Motivation in Managerial Levels: Relationship of Need Satisfaction to Job Performance 312
Smith, Karl U. See Putz, Vernon.
Smith, Karl U., Schremser, Robert, & Putz, Vernon. Binocular Coordination in Reading 251
Smith, Patricia Cain. See Wollack, Stephen.
Smith, Patricia Cain. See Zedeck, Sheldon.
Stone, Thomas H. Effects of Mode of Organization and Feedback Level on Creative Task Groups 324
Streeter, Harrison. See Williges, Robert C.
Taub, Harvey A., & Abrams, Charles. Effects of Target Value and Exposure Duration on Recall in a Visual Search Task 393
Thompson, Duane E. Favorable Self-Perception, Perceived Supervisory Style, and Job Satisfaction 349
Tiffin, Joseph. See Raubenheimer, I. van W.
Triandis, Harry C. See Fiedler, Fred E.
Turner, John. See Rosen, Hjalmar.

Uhes, Michael J., & Shybut, John. Personal Orientation Inventory as a Predictor of Success in Peace Corps Training 498
Vale, Carol A. See Zedeck, Sheldon.
Valenzi E. R., & Andrews, I. R. Effect of Hourly Overpay and Underpay Inequity when Tested with a New Induction Procedure 22
Valenzi, E. R., & Andrews, I. R. Effect of Price Information on Product Quality Ratings 87
Vernon, David T. A. Information Seeking in a Natural Stress Situation 359
Vidaček, Stjepan, & Wishner, Julius. Influence of Expectation of Task Duration on Efficiency of Muscular Activity 564
Vroom, Victor H., & Pahl, Bernd. Relationship between Age and Risk Taking among Managers .. 399
Waters, L. K. See Hulin, Charles L.
Waters, L. K., & Roach, Darrell. Relationship between Job Attitudes and Two Forms of Withdrawal from the Work Situation 92
Weitz, Joseph. See Leonard, Skipton.
Weitz, Joseph. See Levine, Edward L.
Wicker, Allan W., & Mehler, Anne. Assimilation of New Members in a Large and a Small Church. 151
Wijting, Jan P. See Wollack, Stephen.
Williges, Robert C., & Streeter, Harrison. Display Characteristics in Inspection Tasks 123
Wilson, Stephen R. Leadership, Participation, and Self-Orientation in Observed and Nonobserved Groups 433
Wilson, Lowell T. See Searls, Donald J.
Wishner, Julius. See Vidaček, Stjepan.
Wolfe, John H. See Abrahams, Norman M.
Wollack, Stephen, Goodale, James G., Wijting, Jan P., & Smith, Patricia Cain. Development of the Survey of Work Values 331
Wood, Donald A. Background Characteristics and Work Values Distinguishing Satisfaction Levels among Engineers 537
Wright, Patricia. Linguistic Description of Auditory Signals 244
Zedeck, Sheldon. Identification of Moderator Variables by Discriminant Analysis in a Multipredictable Group Validation Model 364
Zedeck, Sheldon, Cranny, C. J., Vale, Carol A., & Smith, Patricia Cain. Comparison of "Joint" Moderators" in Three Prediction Techniques 234

MONOGRAPH

Hackman, J. Richard, & Lawler, Edward E., III. Employee Reactions to Job Characteristics 259
Lawler, Edward E., III. See Hackman, J. Richard.

Journal of Applied Psychology
1971, Vol. 55, No. 1, 1–2

EDITORIAL

It seems appropriate for an incoming Editor to share some initial reflections with the Journal's readers and authors. The *Journal of Applied Psychology* was founded in 1917 by G. Stanley Hall, who served as its Editor until 1920. Subsequent Editors have included James Porter (1921–1942), Donald Paterson (1943–1954), John Darley (1955–1960), and Kenneth E. Clark (1961–1970).[1] With the present volume the Journal has a new Editor for only the sixth time in its 53-year history! My first reflection, then, is on the long tradition of the Journal and on my responsibility to maintain the historical continuity, while encouraging an editorial policy tuned to the times and open to new departures.

At this writing, I have already spent 10 months processing manuscripts for the current volume. During this difficult "breaking in" period I have tried to formulate some tentative ideas about journal policy. These ideas might give some clues as to the kinds of manuscripts we would like to encourage. The *Journal of Applied Psychology* has been the primary outlet, published by the American Psychological Association, for research in various fields of personnel and organizational psychology, as well as engineering and consumer psychology. This tradition will continue. It is our intention, however, to encourage a broadened base of articles in applied psychology with increasing emphasis upon research on socially relevant problems.

The times have generated critical questions that need to be answered by sound research. Many of the problems and concepts of the real world can be translated into psychological language and put to empirical test. Much of what was formerly the province of conventional wisdom, traditional morality, or political action can be dealt with by the psychologist with his quantitative empiricism and from his theoretical base. Similarly, there is a need for validation of concepts and generalizations developed in the laboratory or in hypothetical situations, or with college sophomores, against the hard social realities which surround us. Can we apply the research knowledge and methods already developed toward the solution of pressing problems? Can we be sufficiently innovative to develop new methods for dealing with these research problems, often in an action-oriented setting, with sufficient scientific rigor to allow dependable generalizations? We have heard the challenge from many quarters. Should not the *Journal of Applied Psychology* be a primary outlet for research addressed to these questions?

There are already signs that many of the traditional methods of industrial psychology are increasingly being applied in a wider arena of social problems. Research on selection, training, criterion development, job attitudes, group influences, motivation, communication, organizational behavior, working conditions—all of these have relevance to today's issues. A modern *Journal of Applied Psychology* can contribute by encouraging publication of research in a wide variety of settings in our society. In addition to industry, such settings would include problems of urban areas, criminal justice systems, health systems and institutions, transportation and defense systems, space and other new environments, educational systems, etc. If the Journal is truly to concern itself with psychological research and knowledge applied to life, we need to bring various psychological disciplines into a more cohesive relationship moving toward the understanding and solution of applied problems in these settings.

Examples of work which would be relevant to these pages would surely include: methods of upgrading the disadvantaged in employment and educational settings; problems of test bias with minority groups; police selection, training, and community relations; job restructuring, including improved utilization of paraprofessionals; changing occupational roles of women; adaptation to technological innovation; selection and training for, and adaptation to jobs in foreign cultures (including Peace Corps assignments); studies of value conflict between youth and current institutions; problems of stress in jobs with sensory underload (or overload); longitudinal studies of planned and unplanned change and the effects of intervention programs; factors related to more effective decision making in a variety of situations.

At the same time there is a need for an integrated body of generalized, verifiable, substantive knowledge. We will favor conceptually based research—studies which emerge from concepts or theories which have been or are being developed, particularly if the theory appears relevant to applied issues. We want to avoid an overemphasis on techniques and on articles showing that what we have done well before, we can do again with still another sample. Much of what gets submitted provides no *explanation* of what has been learned to serve as a basis for generalization.

---

[1] Darley (1968) has provided an excellent review of the Journal's first 50 years. During the first two years of the Journal, L. R. Geissler and John W. Baird served as co-editors with Hall. William F. Book served as co-editor with Porter from 1921 until 1927.

We would encourage authors *not* to send in articles prematurely on some fragment or limited aspect of their research. Rather, we feel the *Journal of Applied Psychology* should encourage articles describing more comprehensive research, in some cases organized statements resulting from several experiments. This may mean some longer articles, but the author is cautioned that concise writing is still required with introduction, literature references, and discussions sharply focused on the main points of the article. In the competition for precious journal space, each manuscript will be evaluated on four primary criteria: (*a*) significance in contributing new knowledge to the field, (*b*) technical adequacy, (*c*) appropriateness for the *Journal of Applied Psychology*, (*d*) clarity of presentation. We want to discourage articles of limited interest. We will not publish simple test validation studies, for example, unless some new concept or important methodological or substantive development is reported; nor is a study of the best hammer size for a given size nail likely to excite us.

We recognize the communication sharing function of the Journal and expect to have a *Short Notes* section for brief reports, such as replication studies, methodological notes, presentation of ideas or theoretical discussion, especially interesting pilot studies, etc. These will normally be confined to two journal pages.

I realize the above is a highly idealized statement of policy. As Darley (1968) has commented, "Editors may propose, but authors and the availability of manuscripts dispose." An APA journal is representative of the research of APA members rather than of the aspirations of its editors. Authors and readers have justifiable expectations which place constraints on new departures. Any change is likely to be a gradual process and must reflect the changing nature of the field and many other social forces. We have tried to anticipate some of these.

The reader may be interested in our experience with manuscripts submitted during the first 10 months of 1970. The 286 manuscripts submitted during this period run heavily toward the area of job satisfaction–motivation, with one in every five manuscripts falling into this category. The second largest category of submissions has been in the area of human performance and engineering psychology, followed by selection studies, organizational studies, and research on group processes, consumer psychology, and management.

The number of manuscripts submitted by the end of the year will undoubtedly exceed the total of 311 submitted last year. The acceptance rate is running about 27%. Since this includes acceptances made provisional on extensive revisions, it is likely that the "actual" acceptance rate will be lower. Of the first 200 manuscripts received, only one was accepted "as is." This underscores our feeling that the Editor should be more than a gatekeeper. Our aim is to give careful editorial feedback on each manuscript, including rejected as well as "accepted with revision" manuscripts. We have been gratified by the expressions of appreciation from our authors, even from those to whom we must regretfully say, "No."

An organized enterprise like a scientific journal is possible because of the cooperation and altruism of a great many people. One of my most pleasant duties as Editor has been to appoint the new Board of Consulting Editors. As the new masthead indicates, we have a distinguished Board and they have been magnificent during this period. All accepted appointment without qualification and have contributed their time and talent generously. The personal and professional pleasure gained in my interactions with this network of reviewers and authors has been one of the rewards of this first year's efforts. The Board was chosen to represent the diverse traditional areas of the Journal as well as the increased scope to which we aspire. We have representation from many different kinds of institutions.

I have also made use of a large number of special reviewers among our colleagues and will continue to do so. Their contributions will be acknowledged during the year.

This is as good an opportunity as I will have to thank my illustrious predecessor, Dr. Kenneth E. Clark for his counsel at the time of my appointment. The Journal owes much to Dr. Clark for his stewardship over the past 10 years and, of course, we will be continuing many of the policies he established. At APA, Helen Orr, Managing Editor of APA publications, and Marlyn Kahn, our Technical Editor, have been especially helpful. Finally, at the American Institutes for Research, I am assisted in this task by Pat Smyke and Sally Stein, who have helped me get this operation organized and launched during this transition period.

EDWIN A. FLEISHMAN

REFERENCE

DARLEY, J. G. A journal is born. *Journal of Applied Psychology*, 1968, **52**, 1–9.

Journal of Applied Psychology
1971, Vol. 55, No. 1, 3–8

DEVELOPMENT OF FIRST-LEVEL BEHAVIORAL JOB CRITERIA

LAWRENCE FOGLI, CHARLES L. HULIN,[1] AND MILTON R. BLOOD[2]

University of California at Berkeley

A procedure for the development of behaviorally anchored rating scales was carried out in order to establish reliable job criteria for grocery clerks. Examples of checker behavior were obtained by interviewing staff personnel of a large Western grocery chain. Dimensions of the job were defined by the researchers, and the grocery staff allocated each behavioral example to the dimension in which it belonged. Items with high agreement of dimension allocation were retained and then rated on a scale of desirability. Final scale reliabilities of items ranged above .97. These procedures provide information with implications for policy development and worker training throughout the organization.

The task of developing reliable and valid job criteria is one of the most challenging and desired objectives in contemporary industrial psychology. Dunnette (1966) has emphasized the need for developing job-behavior descriptions in order to establish relationships between specific job behaviors and global dimensions of success. At the same time, predictor batteries can be developed which are aimed at predicting these specific behaviors rather than global measures (which are normally predicted with a consistent, albeit discouraging, lack of success). The relevance of the specific behaviors to more global criterion measures and predictors should be determined empirically. The present study systematically develops criteria for observing and evaluating the behavior of grocery checkers.

Smith and Kendall's (1963) development of criteria of nursing performance provides a procedure for obtaining dimensions and specific aspects of job behavior based on descriptions of actual behavior. The objective for those authors was to develop rating scales that are equivalent from rater to rater and from occasion to occasion in terms of dimensions and levels of job performance, even though the task was complex and varied. The present study strives to establish criteria scales with high interrater reliability that can be used by raters to judge the magnitudes of attributes of workers.

The goal of such research is to develop synthetic criteria which may be substituted for actual specific behavior with little loss of information. Scales and items must indicate the same quality of performance among raters and among divisions of a company. If a sample of potential raters judged the relative quality of performance indicated by each of a set of items, we would expect that the rank order, the location, and the spread of the items would be highly similar to the same statistics computed for the set of items from another sample of potential raters. Ideally, this means that the regression equation relating the scale values of the items obtained from one sample to the scale values obtained from a second sample would have an intercept of 0.0 and a slope of 1.0. That is, $X_{1i} = X_{2i}$, where X_{1i} is the scale value of the *i*th item in the first sample and X_{2i} is the scale value of the same item computed from the second sample under the constraint that $S^2_{x1i} = S^2_{x2i}$.

METHODS

The initial step was to conduct interviews with staff personnel of a grocery chain concerning the purpose of a grocery checker's job at the same time critical incidents of good–poor behavior of checkers were obtained. The most frequent behaviors were selected and the grocer's terminology was retained.

Three groups of managers, assistant managers, and department heads were selected from three districts of a large grocery store chain in the San Francisco Bay area. The three groups consisted of staff personnel from 10 stores in each district, which differed on the basis of geographical location, as designated by company districts—A, B, C. The content of the items and job description is limited to behavior performed in the checkstand.

---

[1] Now at the University of Illinois, Urbana, Illinois.

[2] Requests for reprints should be sent to Milton R. Blood, Department of Psychology, University of California, Berkeley, California 94720.

One hundred and twenty staff personnel were available for participation in the procedures for the development of rating scale criteria. Interviewing was terminated after a total of 43 interviews for obtaining behavioral examples and critical incidents of checkstand behavior because of redundancy in behavioral incidents in the job domain.

The interviews were structured so that the interviewer recorded the verbatim responses to the structured questions. The questionnaire commenced with the following discussion designed to convince the interviewees of the importance of the study and their cooperation:

> We are making a study of grocery checking and trying to learn in detail just what successful work as a grocery clerk includes. In order to facilitate the performance of grocery clerks for checkout stand procedures, we need to know what aspects of checking performance make an especially effective or ineffective clerk. Managers, assistant managers, and department heads such as yourselves are constantly observing and evaluating the work of the clerk. Therefore, you are in an unusually good position to report effectiveness.

Participation in the study was further assured by convincing the respondents of the usefulness of both the expected results and procedures of the study. Interviewees were informed that the incidents reported would serve as characteristic levels of different dimensions of the job, and that later they would be asked to allocate those behavioral incidents to different dimensions.

Questions about primary and specific aspects of job behavior were similar to the critical incident method of Flanagan (1954). Many respondents gave general behavioral examples, even to such statements as "describe specifically the behavior you would expect from the best checker you have ever known." Global descriptions of "he is an accurate checker," "he is friendly," and "he is a fast checker" are too general and are unacceptable for behaviorally anchored scales. The respondents were able to provide more definite and well thought out examples when the generality of their responses was indicated to them.

TABLE 1

AGREEMENT OF ALLOCATION OF ITEMS TO CHARACTERISTICS

Characteristic	High agreement (Above 60%)	Lower agreement (40%–59%)
Conscientiousness	10	5
Knowledge and judgment	37	13
Skill in human relations	36	5
Skill in operation of register	34	3
Skill in bagging	36	1
Organizational ability of checkstand work	13	11
Skill in monetary transactions	34	0
Observational ability	11	2
Total no. of items	211	40

Among the questions designed to obtain specific behaviors were:

1. Please describe the critical characteristics of the behavior of a grocery checker that make him effective—be specific.

2. Similarly describe the specific behavior of an extremely ineffective checker. Remember incidents in your store that displayed vividly poor performance. It might help you to describe the behavior you would expect from the worst checker you have ever known.

3. Now, think of the last time you saw one of the checkers do something that was very significant in displaying effective performance in the checkstand. How did this action result in increasing the effectiveness of the clerk's overall performance?

After analyzing the verbatim responses to the structured questionnaire, the first two authors categorized the behavioral incidents into eight functional dimensions. Store personnel were not asked to define the job dimensions so that demands on their working time would not be excessive.

Staff personnel were presented with dimensions defined by the researchers and instructed to place each behavioral incident into the dimension for which it was an example. Each behavioral incident was judged by 15 independent raters. An incident was eliminated if there was not majority agreement as to the dimension to which it belonged.

A final rating format composed of all of the items that were maintained in the eight dimensions was sent to all 120 *S*s, and 97 *S*s completed it. Judges rated the degree to which each item indicated good-poor performance on a 7-point scale. Large standard deviations indicated a lack of agreement on the meaning of the behavior, and items were arbitrarily eliminated if their standard deviations were 1.5 or greater. If items had similar mean values, the item with the smallest standard deviation was retained. Items were selected for each categorical scale by these criteria. The mean scale positions of each item assigned by personnel in districts A, B, and C were intercorrelated to give estimates of scale reliabilities. This procedure assumes that the rater can make inferences from observations about the interrelations of behaviors and that a group of managers, assistant managers, and department heads will have well-differentiated and standardized judgments about checking performance.

The behavioral examples that were used in the final rating format were reworded as expected be-

TABLE 2

INTERCORRELATIONS, STANDARD DEVIATIONS, AND REGRESSION DATA OF MEAN SCALE POSITIONS ASSIGNED BY THREE DISTRICTS FOR SEVEN SCALES

Characteristic	No. items	District	Correlations		$\bar{X}$	σ	Regression coefficients[a]			Intercepts[a]		
			A	B			A	B	C	A	B	C
Conscientiousness	13	A			5.89	1.08		1.14	1.03		−.33	.01
		B	.98		5.46	.93	.89		.84	.37		.50
		C	.97	.97	5.71	1.01	.92	1.06		−.28	−.08	
Knowledge and judgment	24	A			4.75	1.74		.97	.90		.30	.66
		B	.98		4.59	1.76	.99		.92	−.11		.43
		C	.99	.99	4.55	1.90	1.08	1.07		−.58	−.36	
Skill in human relations	16	A			4.88	1.57		1.04	.97		.09	.33
		B	.98		4.61	1.48	.93		.90	.08		.35
		C	.99	.98	4.72	1.61	1.01	1.07		−.21	−.21	
Skill in operation of register	26	A			5.04	1.49		1.13	1.04		−.29	.25
		B	.98		4.71	1.32	.86		.92	.36		.30
		C	.98	.98	4.78	1.41	.93	1.05		.12	−.17	
Skill in bagging	21	A			4.27	1.93		1.07	1.01		−.10	−.10
		B	.99		4.09	1.78	.91		.93	.02		.24
		C	.99	.99	4.14	1.90	.98	1.05		−.03	−.14	
Organizational ability of checkstand work	30	A			4.79	1.79		1.07	.95		−.18	.39
		B	.99		4.64	1.65	.92		.88	.25		.22
		C	.99	.99	4.62	1.86	1.03	1.11		−.31	−.53	
Skill in monetary transactions	32	A			4.36	1.72		1.07	.95		−.02	.41
		B	.99		4.09	1.58	1.03		.91	−.40		.31
		C	.98	.99	4.16	1.78	1.01	1.01		−.24	.03	

[a] Regression coefficients and intercepts are given for predicting the variable designated by the row by the variable designated by the column.

haviors rather than actual behaviors. This restatement of the items followed the procedures of the Smith and Kendall (1963) study. Instead of statements like "When everybody is checking, this checker leaves the checkstand to get the customer an item," statements consisted of expectations such as "If everybody is checking, this checker can be expected to leave the checkstand to get the customer an item." According to Smith and Kendall the expected-behavior format will improve the final ratings by making the predictions

> (*a*) so concrete that, in view of previous agreement by the peer group, central tendency or judging effects will be minimized; and (*b*) so verifiable that insight, judgment, values, etc. of the rater are potentially challenged if later behavior of the ratee would fail to confirm the prediction [p. 151].

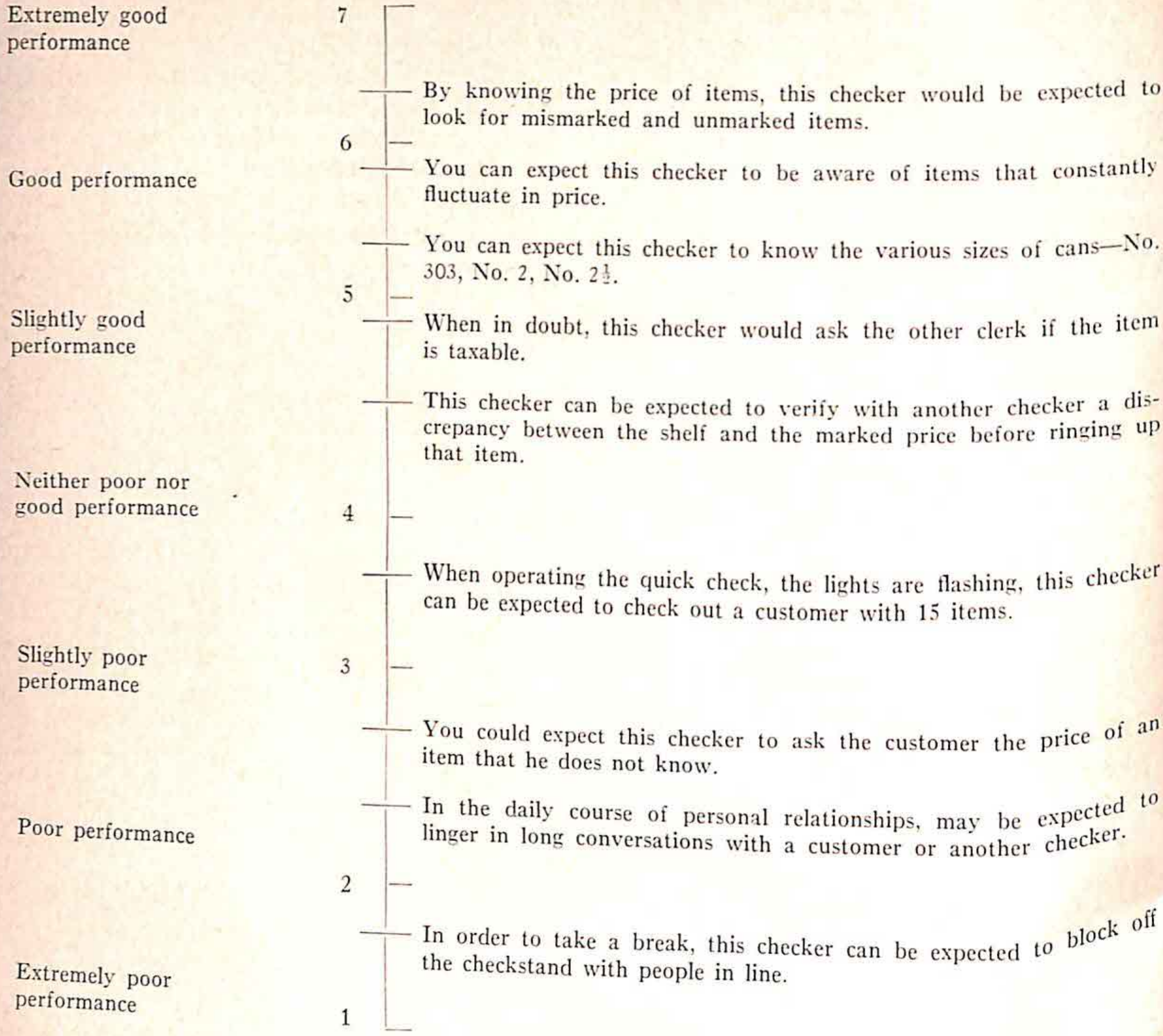
RESULTS

The dimensions that proved to be meaningful after the judges allocated items to dimensions are listed in Table 1. Even though the researchers, rather than job incumbents, defined the dimensions of the job, the total number of items (211) in the high agreement column of Table 1 is indicative of the usefulness of those dimensions. The rating questionnaire was composed of items of at least 60% agreement except for Categories 1, 6, and 8. These categories contained some items with only 50% agreement.

Table 2 shows scale reliabilities of items retained in the final rating format. The mean score of each of the retained items for the three districts was intercorrelated to obtain scale reliabilities.[3] The lowest scale reliability

---

[3] The reader should be aware that pooling item scores leads to more accurate estimates of scale values in this assessment of scale reliability. With


Extremely good performance
7
Good performance
6
Slightly good performance
5
Neither poor nor good performance
4
Slightly poor performance
3
Poor performance
2
Extremely poor performance
1

By knowing the price of items, this checker would be expected to look for mismarked and unmarked items.

You can expect this checker to be aware of items that constantly fluctuate in price.

You can expect this checker to know the various sizes of cans—No. 303, No. 2, No. $2\frac{1}{2}$.

When in doubt, this checker would ask the other clerk if the item is taxable.

This checker can be expected to verify with another checker a discrepancy between the shelf and the marked price before ringing up that item.

When operating the quick check, the lights are flashing, this checker can be expected to check out a customer with 15 items.

You could expect this checker to ask the customer the price of an item that he does not know.

In the daily course of personal relationships, may be expected to linger in long conversations with a customer or another checker.

In order to take a break, this checker can be expected to block off the checkstand with people in line.

FIG. 1. An example scale for the knowledge and judgment dimension.

was .97. Means, standard deviations, and regression information are also given for each district in Table 2. An examination of these data substantiate the similarity of rating between districts on each scale. The regression coefficient and intercept which are required to indicate perfect agreement among districts are a coefficient of 1.00 and an intercept of .00. The observed coefficients range from .84 to 1.14 with a median of .99. The observed intercepts range from −.58 to .66 with a median of .05.

Figures 1 and 2 present example scales made from items which represent two of the dimensions—knowledge and judgment, and organizational ability of checkstand work.[4] Category 8, observational ability, was eliminated as a useful scale because of poor dispersion of behavioral items along the good–poor continuum.

---

30 Ss per sample, this is the empirical equivalent of adjusting the average interrater correlation by the Spearman-Brown formula with a multiplicative factor of 30 (see Noble, 1955, for a discussion of such scale reliability estimation).

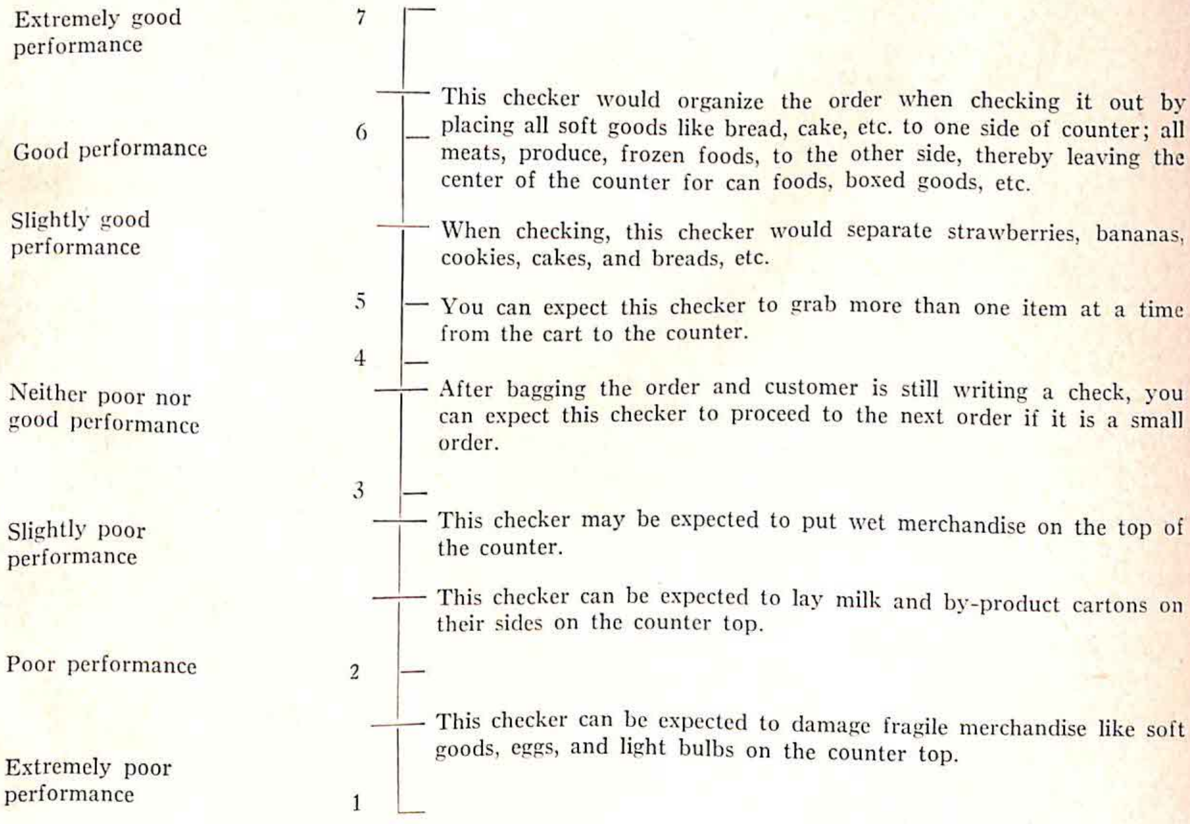
[4] A complete list of items retained for each of the seven scales and the scale values of the items has been deposited with the National Auxiliary Publications Service. Order Document No. 01257 from the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche and make checks payable to Research and Microfilm Publications, Inc.

CONCLUSIONS

The method of criterion scale development reported in this article, which involves staff members in each step of the development, provides behavioral scales that are specific and apparently nonambiguous. In the case of the specific job studied, seven major and discriminable dimensions were obtained: (*a*) conscientiousness, (*b*) knowledge and judgment, (*c*) skill in human relations, (*d*) skill in operation of register, (*e*) skill in bagging, (*f*) organizational ability of checkstand work, and (*g*) skill in monetary transactions. The procedure provides carefully designated behavioral incidents for each scale. The examples of behavior are in the grocer's terminology, and there is excellent agreement of the allocation of incidents, discrimination, and high scale reliability. The degree of colinearity among these scales, when used as rating devices, remains to be determined.

This method eliminates ambiguity by having those who will eventually use the rating criteria develop the scales with their own observations about behavior. Furthermore, the staff allocates incidents to dimensions and makes final judgments about the level of performance each behavior indicates. Since high agreement is obtained in the three samples, the scales may serve as standards to evaluate specific behaviors that can be expected in the checkstand. The reliability of ratings of performance may increase since the dimensions and incidents are easily and accurately distinguished from one another. Because there are multiple ratings, raters are not restricted to the gross generalizations which are necessary when global performance evaluations are made. The success of a job incumbent can be evaluated in terms of specific behaviors and dimensions. This behavioral information should prove more useful to both the worker and his immediate supervisor than general statements of overall performance level.

There are many items with standard deviations less than 1.5 which were eliminated from the scales because their mean scores were similar to other items on the same dimension.


Extremely good performance
7
Good performance
6
Slightly good performance
5
4
Neither poor nor good performance
3
Slightly poor performance
Poor performance
2
Extremely poor performance
1

This checker would organize the order when checking it out by placing all soft goods like bread, cake, etc. to one side of counter; all meats, produce, frozen foods, to the other side, thereby leaving the center of the counter for can foods, boxed goods, etc.

When checking, this checker would separate strawberries, bananas, cookies, cakes, and breads, etc.

You can expect this checker to grab more than one item at a time from the cart to the counter.

After bagging the order and customer is still writing a check, you can expect this checker to proceed to the next order if it is a small order.

This checker may be expected to put wet merchandise on the top of the counter.

This checker can be expected to lay milk and by-product cartons on their sides on the counter top.

This checker can be expected to damage fragile merchandise like soft goods, eggs, and light bulbs on the counter top.

FIG. 2. An example scale for the organizational ability of checkstand work dimension.

These items still provide reliable job criteria. This company could use these items plus the behavioral scales for development of a training program for checkers. A training program based upon the behavioral incidents and dimensions would inform checkers about the expected behaviors for their job. Since the level of the behaviors is highly agreed upon from store to store, a uniform training program could be developed for all of the stores.

An additional use of the behavioral items would be to ask trainees and/or incumbent checkers to rate the items on the 7-point good–poor scale. Their ratings of items could be compared to the ratings made by managers. Discrepancies could be discussed to teach the checker what he should do. Training programs could be modified to take account of the results of such item ratings.

Another benefit of this procedure is that it provides information about specific behaviors that are not highly agreed upon among staff personnel. Items with very large variances can be investigated by the organization to change or develop policy about such behavior. The following are examples of behavioral incidents which had large variances on the good–poor continuum in this study: (*a*) When customer feels that the total bill was too high, you can expect this checker to rering the order ($S^2 = 3.20$). (*b*) When dealing with an overring, this checker would subtract overcharged items from another item in the same department—a grocery item from a grocery item ($S^2 = 3.02$). (*c*) You can expect this checker not to separate wet produce from other merchandise on the table top ($S^2 = 2.53$). (*d*) You can expect this checker to check items without looking at the prices because of memorizing prices ($S^2 = 3.50$). (*e*) This checker can be expected not to bother to change departmental sales ring-up error but would ignore the mistake ($S^2 = 2.43$). For these behaviors and others with large variances, the organization may wish to enunciate or communicate its policy more clearly.

REFERENCES

Dunnette, M. D. *Personnel selection and placement.* Belmont, Cal.: Wadsworth, 1966.

Flanagan, J. C. The critical incident technique. *Psychological Bulletin,* 1954, **51**, 327–355.

Noble, C. E. Scale reliability and the Spearman-Brown equation. *Educational Psychological Measurement,* 1955, **15**, 195–205.

Smith, P. C., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology,* 1963, **47**, 149–155.

(Received February 2, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 9–21

HUMAN FACTORS IN OPERATING SYSTEMS RELATED TO DELAY AND DISPLACEMENT OF RETINAL FEEDBACK[1]

VERNON PUTZ AND KARL U. SMITH[2]

Behavioral Cybernetics Laboratory, University of Wisconsin

A central question of the theory of human factors in man-machine relationships concerns the way in which operators correct for and adapt to spatial noncompliances in feedback between eye, head and manual movements, and visual input. Therefore, the nature of response to displaced and delayed vision represents also a theoretical issue for human factors theory of machine operation. Theories of response to altered retinal feedback—i.e., associative learning doctrines and the feedback-compensation hypothesis—were evaluated in relation to their application in defining human factors principles in machine and perceptual training designs. Using 12 *S*s, controlled comparisons were made of the relative effects of reversed and delayed feedback of head and eye movements under conditions in which head movements could not compensate altered feedback of eye movements and vice versa. Accuracy of ocular tracking was affected more by reversal of retinal feedback of eye movements than head tracking was changed by reversal of visual effects of head movements. Eye tracking was reduced to highly irregular, unstable vision, whereas saccadic head movements retained a relatively high level of tracking accuracy. Head-eye tracking combined the effects of reversal of feedback that were found with each of the specific types of movement. Delay of retinal feedback of the movement systems at intervals of 0.0, 0.2, 0.4, 0.6, 0.8, and 1.5 sec. caused progressive elimination of unstable ocular guidance, but also greatly increased the error of such tracking. Other results indicated that there was little or no learned adaptation to the reversed and delayed vision produced by head and eye movements. The findings give support to a behavioral cybernetic interpretation of the guidance factors in man-machine and perceptual systems relationships by showing that the effects of altered feedback in machine and systems operation are determined by movement capabilities in compensating displacements and delays in sensory input. The results also suggest that visual impairments may be produced by delays in the retinal feedback effects of eye and head movements and that these defects may require dynamic methods of optometric diagnosis and training for their measurement and correction.

This experiment assesses the validity of the theories of adaptation to displaced and delayed retinal feedback in behavior relative to the human factors problems of design of machines and tools, and of optometric diagnostic and visual training and rehabilitation devices. Although the events of optical distortion and displaced vision have been studied extensively in the attempt to justify the learning of space perception (Freedman, 1968), the relevance of the results of such research to the human factors problems of design and operation of devices and perceptual measurement and training systems has not been made clear. The purpose of this research is to evaluate the validity of past theory of perceptual adaptation in relation to a behavioral interpretation of displaced and delayed vision and to suggest applications of the more accurate view to current theory of human systems design.

In the behavioral feedback approach to be defined here, the core experimental problem of displaced and delayed vision concerns the relative effect of such feedback alteration on head and eye movements. The theoretical issue is whether adaptation to displacement and

---

[1] This research was supported by grants from the National Science Foundation and the Social and Rehabilitation Administration, and was carried out in the context of a training program in psychophysiological cybernetics supported by the Biological Training Section of the National Institute of Mental Health.

This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Karl U. Smith, Behavioral Cybernetics Laboratory, University of Wisconsin, 918 Conklin Court, Madison, Wisconsin 53715.

delay of retinal feedback of these two movement mechanisms is determined by movement compensation of the altered feedback, or the result of learned association between visual and tactual experience. To test this specific question, the present experiment investigated the problem of response to reversal of the retinal image in a new way by dynamically isolating the retinal effects of delay and space displacement of eye and head movements. The purpose of the experiment was twofold: (*a*) to test the feasibility of development and use of controlled systems methods in the measurement and diagnosis of visual distortions; and (*b*) to measure the effects of retinal feedback delay and distortion on eye and head movements under controlled conditions in which it was not possible for eye movements to compensate feedback alterations of head movements, and vice versa. The methods aspect of the study has implications for optometry and perceptual training by defining certain experimental foundations for dynamic feedback methods of optometric measurement and visual training in perceptual disorders.

As noted above, two general theories have been formulated regarding adaptation to altered visual feedback: the classical theory of von Helmholtz (1856–66) and the movement-compensation hypothesis of Smith and Smith (1962) and of Hay and Pick (1966). According to the classical view and its modern restatement (Freedman, 1968; Held & Hein, 1958; Rock & Harris, 1967), normal space perception and adaptation to optical distortions is dependent on learned association between visual and tactual (or proprioceptive) processes. Proponents of this view assume that individuals can adapt more or less completely to different conditions of displaced vision and that the adaptive change consists of a general process of perceptual learning. The movement-compensation hypothesis asserts that the capability of response and adaptation to displaced vision depends specifically on the movement mechanisms—body, limb, head, or eye—affected by altered feedback and the capacity of a second movement system to compensate the one that is affected by the visual distortion. Smith and Smith provided specific evidence for this hypothesis by showing that the extent of adaptation to displaced vision varied with the type of motion affected and the conditions of displacement. Hay (1968) showed that adaptation to altered feedback of head movements varied in relation to the degree that the eyes could correct for the head movement feedback displacement. Wallach, Kravitz, and Lindauer (1963) showed also that rapid adaptation to displaced visual direction depended on the mode of movement and position of *S* relative to the altered spatial environment. The present study extends these past investigations by developing controlled systems methods for isolating the sensory effects of delayed and displaced retinal feedback of head and eye motions.

METHOD

The experimental vision station and hybrid-computer system diagramed in Figure 1 was used to yoke eye, head, and coordinate head-eye movements to a light target that was projected on a cylindrical viewing screen located 36 in. from *S*'s eyes. The *S*'s task consisted of moving his eyes, head, or head and eyes together to keep the target centered on the screen, as it was displaced horizontally by the computer system. In addition, the computer system was programmed to reverse the spatial relation between movement direction and target direction, and as a time machine to delay the action of the yoked feedback signals of the different movement systems for predetermined periods of time.

The insert in the upper right in Figure 1 illustrates the techniques of transducing the eye, head, and head-eye motions. Eye movements were transduced by the electrooculographic method (Mowrer, Ruch, & Miller, 1936; Smith, Kappauf, & Bojar, 1940), while head rotation in the horizontal plane was recorded by a headgear that rotated a potentiometer located at the top center of rotation of the headgear. The two transducing systems were approximately linear in their output and could be exactly equated electrically by the calibration procedures used.

The experiment was controlled and feedback-delay variations made in the yoked eye and head movements by the computer system. To effect this control, the continuous analog signals of the transduced movements were amplified and converted to digital form by the analog-digital converter. This converter also measured the incoming signals on a nine-bit digital voltage scale. The digitized signals were then selected by the digital computer in which they were programmed for calibration of the transducing systems, for experimental measurement, and for introduction of delays into the feedback outputs of the systems. After these programmed operations were completed, the signals were reconverted to analog

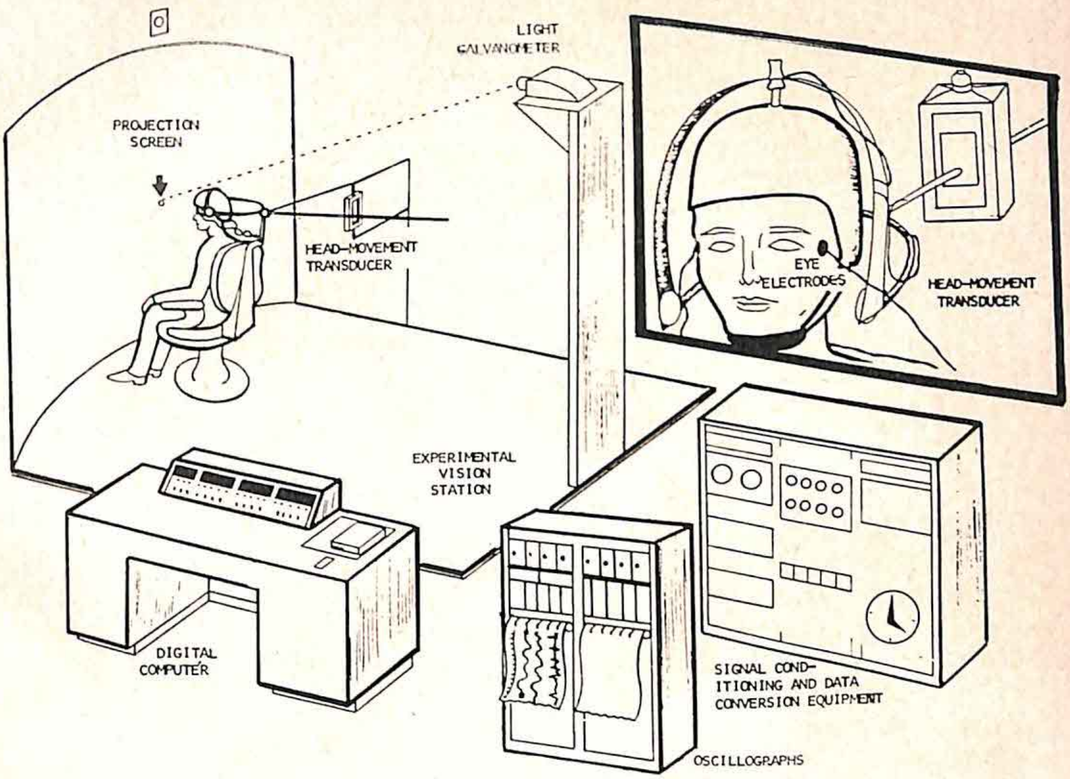


FIG. 1. Experimental vision station and hybrid computer laboratory system for yoking projected visual targets to eye and head motions. (Insert: Eye movement and head movement transducers used in the experiment.)

form by a digital-analog converter. Thereafter, they were used to activate a light galvanometer which threw a light spot of 1.5° in diameter on the projection screen.

Details of hybrid computer control and calibration of the transducers of the experiment are illustrated in Figure 2. To measure head and eye tracking error, the zero point of the screen was established as a zero voltage value and voltage variations from that zero point, as produced by the transducers, measured by the computer system. This measured error was outputted to control movement of the target which *S* saw. Before each trial was run, *S* was instructed to view the center point of the screen and the computer established this fixation position of the particular transducer as a zero value.

Both the eye- and head-motion signals were subjected to a second calibration between experimental trials. As indicated in Figure 2a, targets of known deviation were projected on the screen, and *S* was required to fixate these targets by head or eye movement. A deviation from equivalence between target-voltage value and transducer-voltage value at the fixated positions was corrected by increasing or decreasing the eye and/or head transducer signal with a pair of potentiometers which were wired into the control circuits. The relations between this linear function relating calibrated transducer position with respect to target position, oscillograph pen deflection, and tracking error magnitude are given in Figure 2b.

The sequence of automated computer control of different experimental operations is illustrated by the oscillograph record of tracking error shown in Figure 2c. In tracking the yoked light spot on the screen, *S* observed all the phases of operation depicted by this record. The error record, which represented 55 sec. of performance, was seen as a deviation of the light spot from the center of the screen. The black marks, which represented warnings to *S*s to fixate the center of the screen, were seen as rapid 5° oscillations of the projected light target. The specific operations labeled were all controlled as phases of experimental automation that were regulated by dynamic computer programming.

In viewing the target on the screen during an experimental period, *S* saw a light spot about 1 in. in diameter which deviated from the center zero position. By moving his head or eyes in a direction opposite this deviation, the target could be brought back to center. Error voltage was measured as a deviation in making this compensatory centering movement. To cause the target to deviate in the manner just indicated, the computer was programmed to induce a random sine-wave variation in target position which had a mean frequency of 4.31 cpm. Also, in different trials, the computer was programmed to hold incoming digitized eye- or head-motion signals in memory for specified periods o

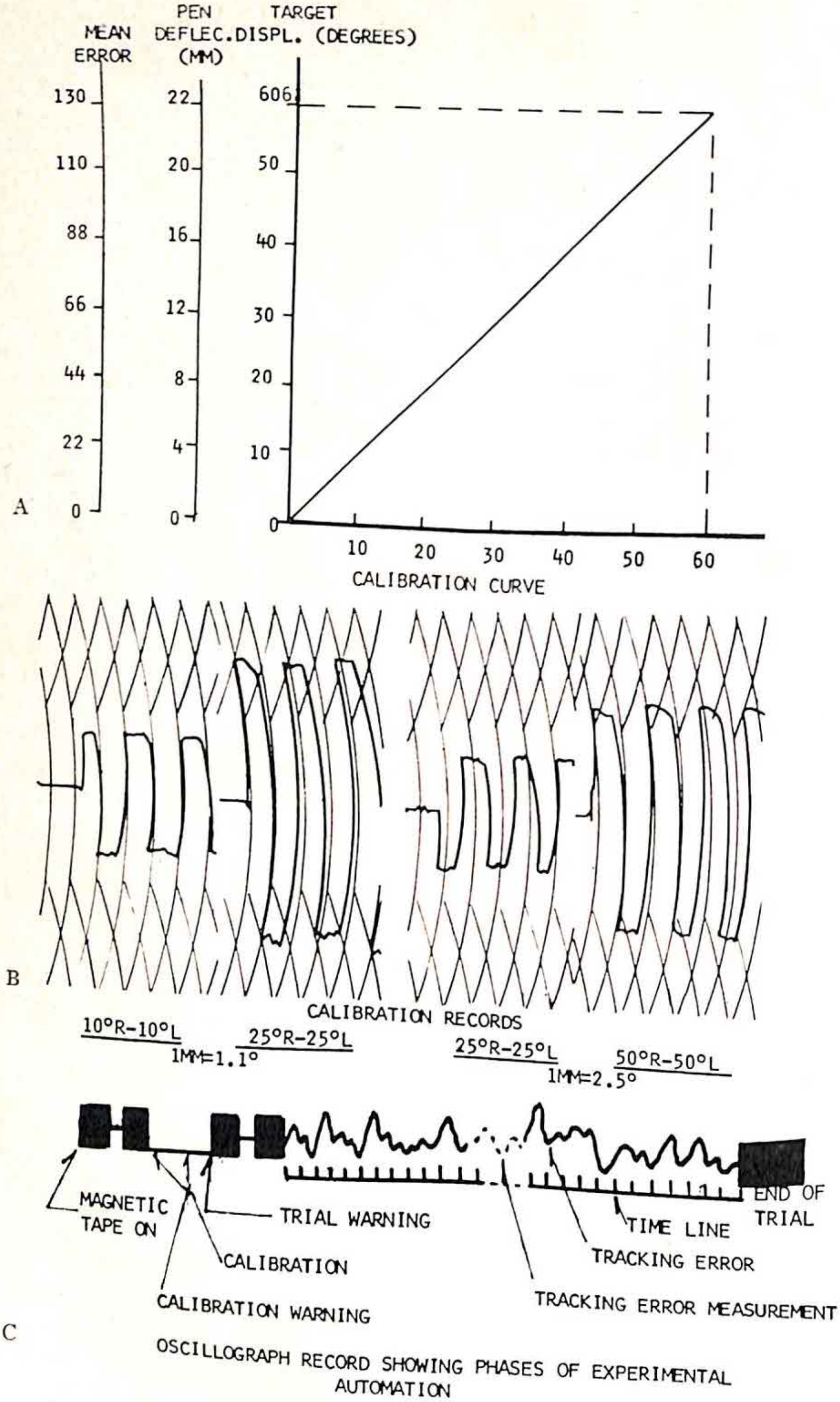


Fig. 2. Dynamic programming techniques used in experimental calibration (a and b) and experimental automation (c). ([a] Curve showing how target voltage and eye-movement voltage were equated and related to measured tracking error and oscillograph pen displacement. [b] Calibration records showing the linearity of response of the control and recording system. [c] Steps in experimental automation by dynamic programming as indicated by an oscillograph record of eye-tracking error.)

0.0, 0.2, 0.4, 0.6, 0.8, and 1.5 sec. before outputting these signals for delayed feedback control of the yoked target. Under the delayed feedback conditions, S experienced a lag effect in action of the target with respect to his head or eye movements which was observed as an optical error or displacement of the target. In the course of observations, the computer automatically measured the tracking error for each selected sample of the digitized input movement signal—that is, .128 times per second. These sample error measures were reduced on-line to 5-sec. error means.

A repeated-measures plan with five factors was used in which each of 12 Ss acted as his own control in performing the three modes of tracking with six delay intervals. The design was three modes of movement or tracking of 18 trials each with each S, and nested the six delay magnitudes of 0.0, 0.2, 0.4, 0.6, 0.8, and 1.5 sec. within each mode of tracking. Conditions of delay were ordered according to a counterbalancing scheme. This scheme resulted in three trials of 55 sec. each for each delay. Within each trial there were ten 5-sec. error measurement periods. Between successive trials, Ss were rested for 35 sec. and for periods of 5 min. after 15 min. of observation. Periods of observation lasted for 2 hr.

The crux of experimental control in this study concerned the isolation of head and eye motion under particular modes of head, eye, and head-eye tracking. Under a particular mode of tracking, only the transducer or transducers peculiar to that mode would actuate the yoked target. In eye tracking, S's head was held fixed, he was instructed to move only his eyes, and was informed that any head movement would only increase error. In head tracking, S was instructed to keep his eyes fixated on the center point of the projection screen, and was told that any use of the eyes in tracking would serve only to increase error. In coordinated head-eye tracking, both the head and eye transducers were activated and S could control the yoked target with either head or eye movement equivalently.

RESULTS

The main results consist of comparisons of tracking accuracy between the three movement modes and between the different delay intervals.

Effects of Retinal Feedback Reversal and Delay

The three movement records shown in Figure 3 will tell generally what happened in the experiment with the three modes of tracking control. The records shown are of the actual movements that the eye or head performed in tracking a computer-generated sine-wave target variation like that shown to the left of the records. To track accurately, the eye or head had to oppose or compensate this wave pattern to keep the target centered on the projection screen. As the records show, the head performed the smoothest tracking, the eye movements were far less regular and accurate, and the head-eye motions were least

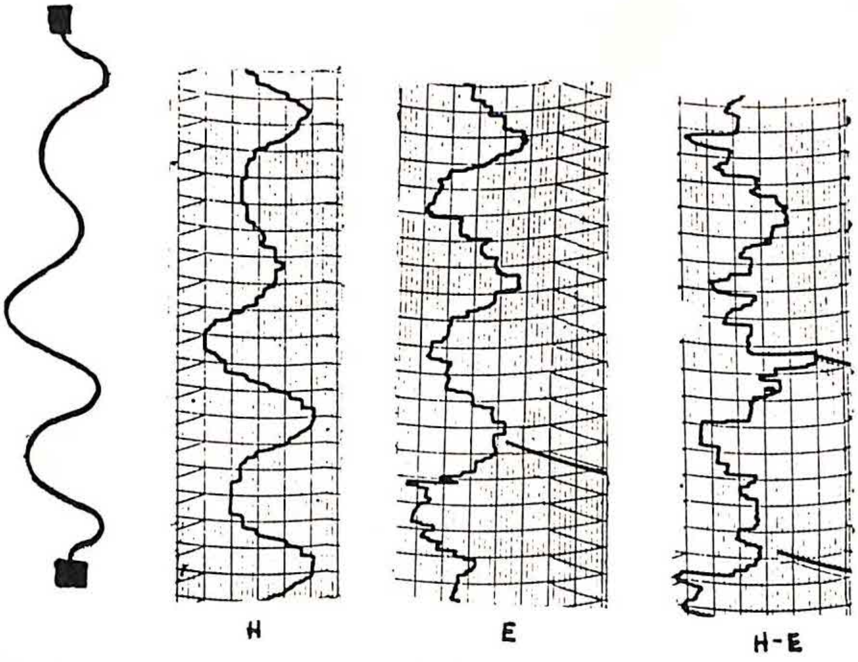


FIG. 3. Oscillograph movement records from one S with reversed retinal feedback with the three modes of tracking.

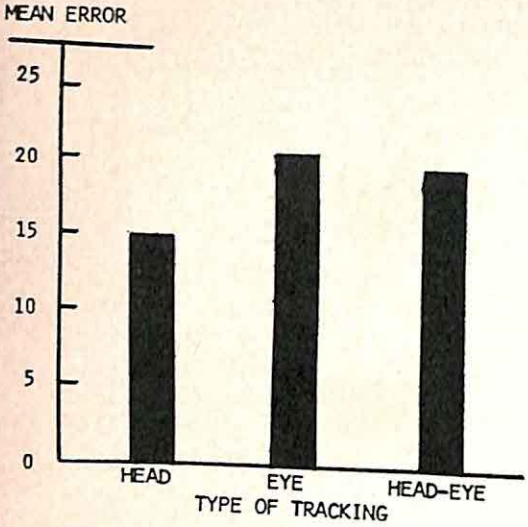


FIG. 4. Bar graph showing the differences in accuracy of eye, head, and head-eye tracking.

accurate. In the case of all three movement modes, tracking motions, instead of being smooth pendular pursuit movements as they should have been, were step-wise saccadic motions. These were smallest for head movements, which tracked the target most accurately.

The bar graph in Figure 4 describes the relative error of the compensatory eye, head, and head-eye tracking for all experimental conditions combined. The graph indicates that head control gave the most accurate tracking, whereas accuracy of the eye and head-eye tracking were both considerably lower than the precision of head movements.

The functions relating tracking error in the three movement modes with feedback delay magnitude are given in Figure 5. The superior accuracy of head tracking was most prominent with lower magnitudes of delay, but was reduced progressively as the feedback delay interval was lengthened. Accuracy of eye and head-eye tracking was almost identical at all magnitudes of feedback delay. For all three movement modes, the effects of eye-retinal and head-retinal transmission lags were very nearly linear.

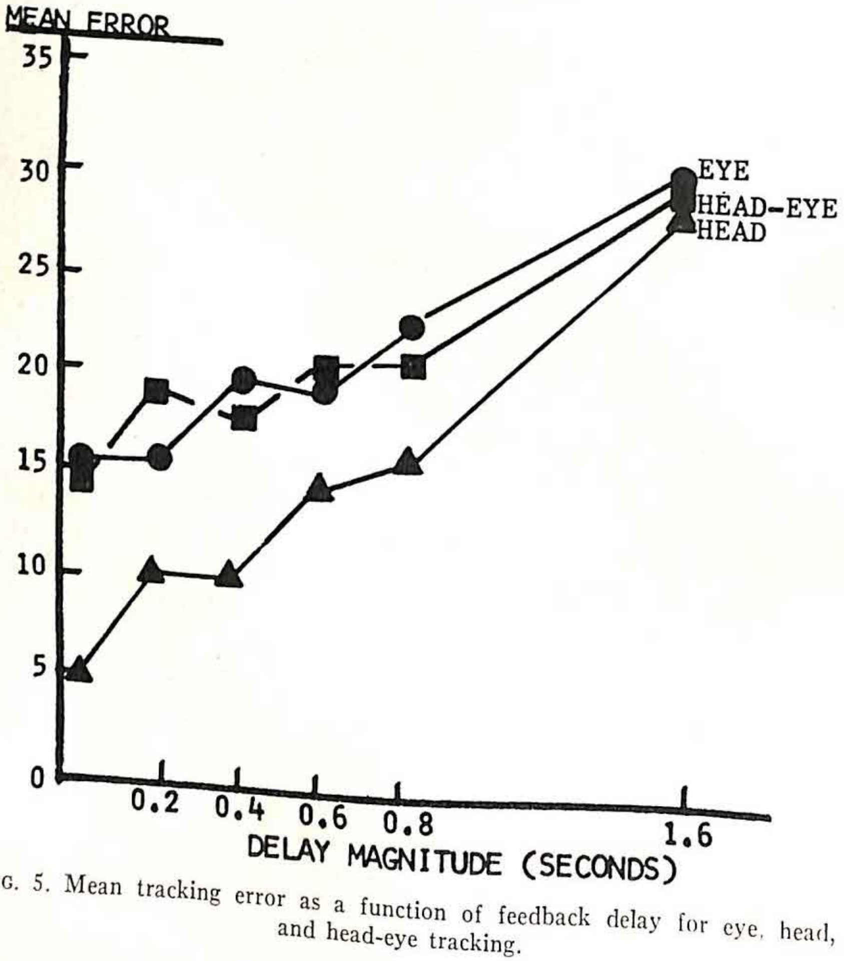


FIG. 5. Mean tracking error as a function of feedback delay for eye, head, and head-eye tracking.

Head and Eye Adaptation to Reversed and Delayed Feedback

Characteristics of adaptation of eye and head motions to the reversed and delayed feedback were analyzed in terms of oscillograph records of tracking movements and tracking errors under the variable feedback conditions. Records of head, eye, and head-eye movements of three *S*s are shown in Figure 6. Each series of records represents performance in one trial at each of the delay intervals. Aside from the seconds time line given to the left, each record contains three lines repre-

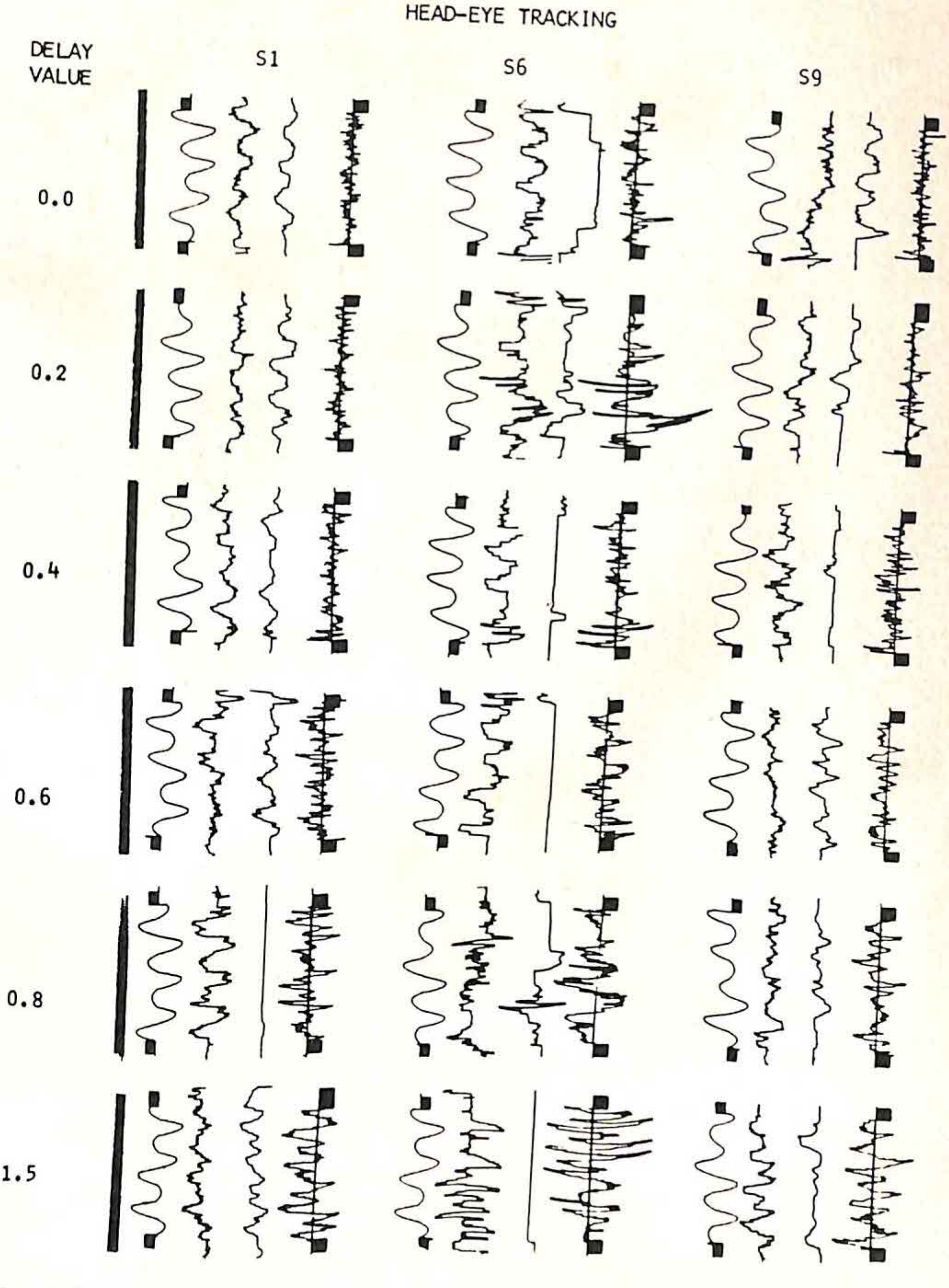


FIG. 6. Oscillograph records of head, eye, and head-eye tracking of three *S*s in one trial at the six retinal feedback-delay intervals.

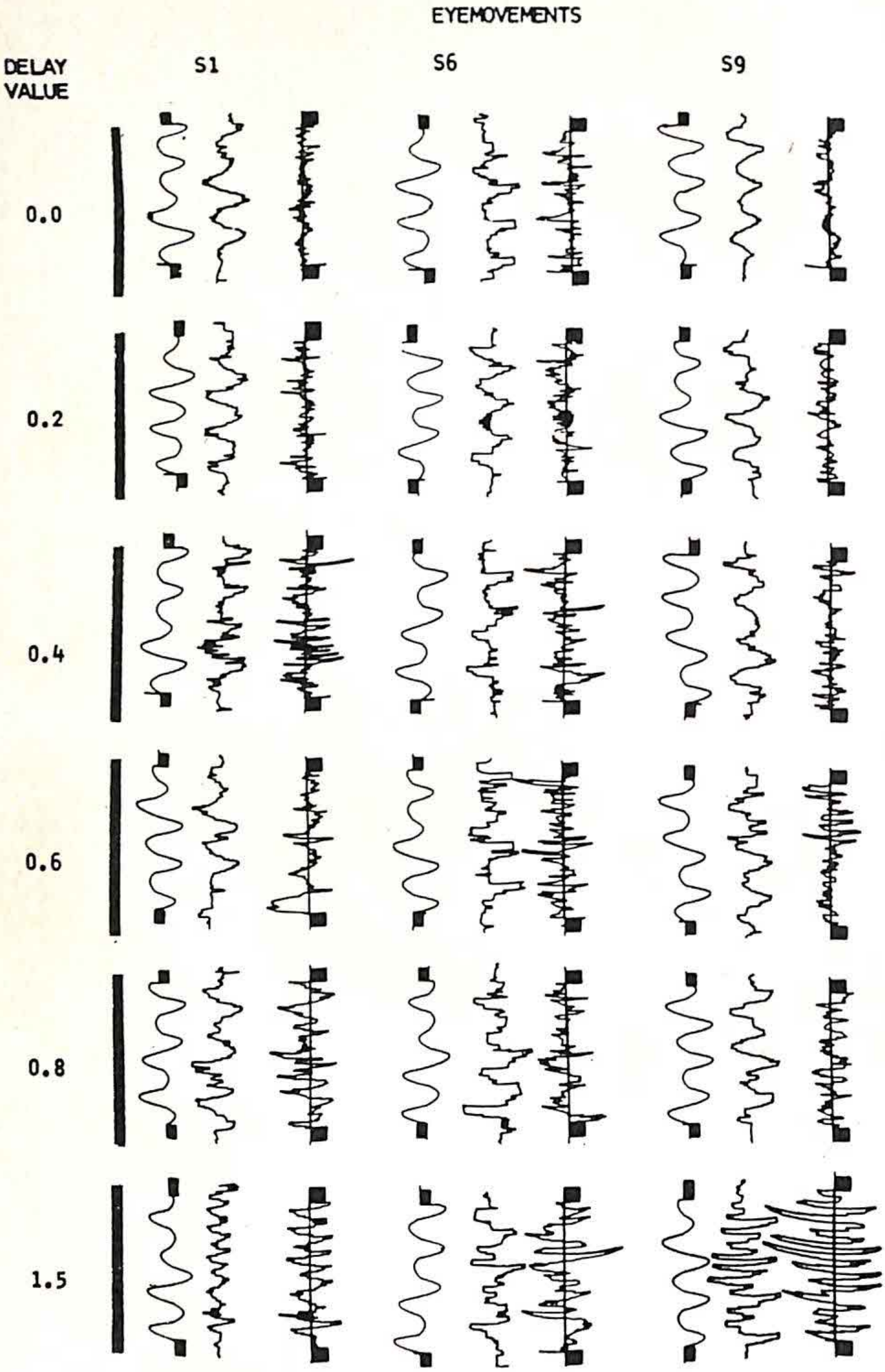


FIG. 6. Continued.

senting the target wave, head or eye movement, and tracking error (to the right). In the case of the coordinate head-eye movements, each record has four lines representing target wave, eye movements, head movements, and coordinate head-eye tracking error.

Overlooking some variations in the magnitude of tracking error, the three *S*s showed quite similar head, eye, and coordinate head-eye movements under the six delay conditions. At zero delay, the head movements were very accurate, composed of fine saccadic move-

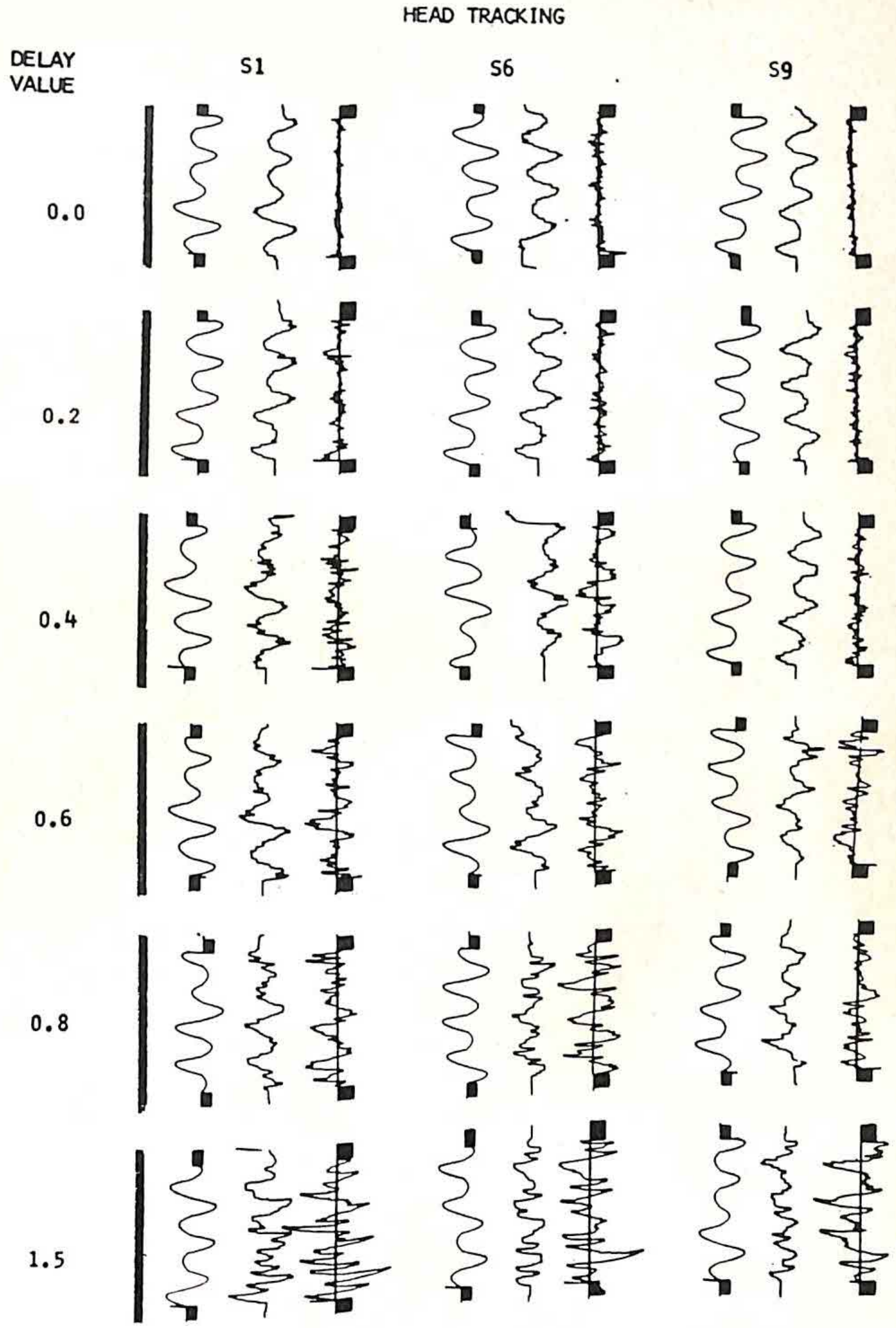


Fig. 6. Continued.

ments, and duplicated the pattern of the target wave. In contrast, eye movements at zero delay were highly erratic although also composed of saccadic movements. Under these same conditions of reversed feedback at zero delay, head-eye movements were also highly erratic; S_6 did almost all the tracking in this situation with his eyes.

The records showed that with increasing magnitudes of retinal feedback delay, tracking error increased progressively. This increased error was manifest especially in

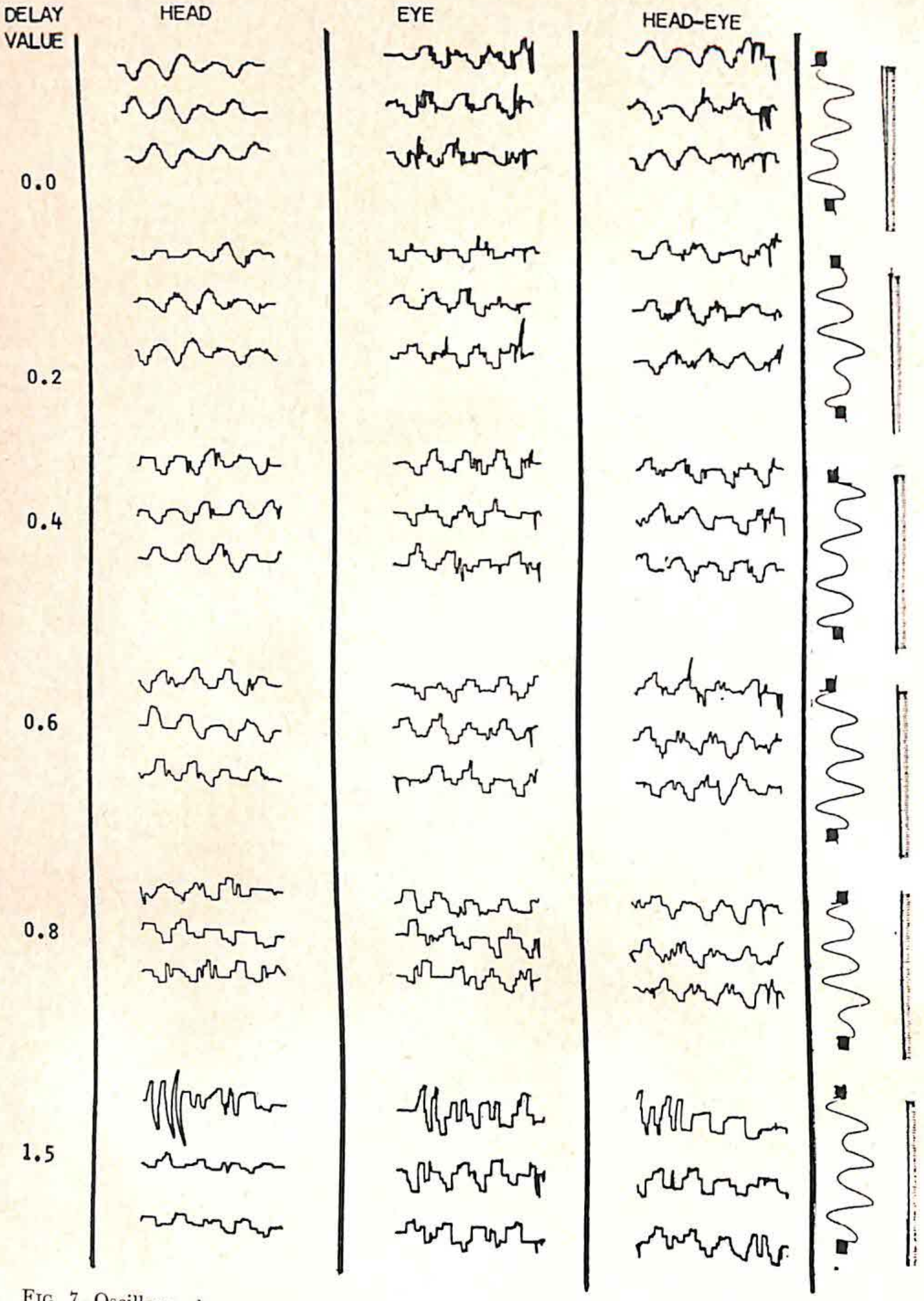


FIG. 7. Oscillograph records of head, eye, and head-eye movement for one *S* at each of the three trials with all six intervals of feedback delay.

marked change in irregularity and magnitude of eye movements. In the case of head-eye movements, head motions became less prominent in the coordinate action in all three *S*s as the delay interval was increased. This was especially true of S_6 and S_9.

An overall summary of all the records of the head, eye, and head-eye movements o one *S* can give an indication of the extent to which adaptation occurred with practice a the particular delay intervals. Such a summary is given in Figure 7. Each of the sets o records in the figure represent the three successive trials of each tracking mode at a par

TABLE 1

DUNCAN MULTIPLE-RANGE TESTS FOR DIFFERENCES IN MEAN TRACKING ERROR FOR MOVEMENT MODES IN EYE, HEAD, AND HEAD-EYE TRACKING

Level of significance	Head	Eye	Head-eye
5%	14.80	20.40	19.82
1%	14.80	20.40	19.82

Note.—Lines underlying common means are not significantly different from one another.

TABLE 2

DUNCAN MULTIPLE-RANGE TESTS FOR DIFFERENCES IN MEAN TRACKING ERROR FOR FEEDBACK-DELAY INTERVALS

Level of significance	0.0	0.2	0.4	0.6	0.8	1.5
5%	11.60	15.32	16.00	17.51	19.70	29.96
1%	11.60	15.32	16.00	17.51	19.70	29.96

Note.—Lines underlying common means are not significantly different from one another.

ticular delay interval. Some sequences of trials, such as those at 1.5-sec. delay, show some indications of improvement over the three trials, while other sequences, such as those for the head and eye at 0.6-sec. delay, give evidence of increased irregularity with practice. Overall, the changes with practice in the head, eye, and head-eye movements were indeed quite limited.

Statistical Analysis

A five-way analysis of variance was carried out on the error data. This analysis was based on the 5-sec. error means within trials or observations, and encompassed the means for trials, delays, movement modes, and *S*s. All the main variables were significant at the 1.0% level, as tested by the appropriate error terms. The trial variable, which indicated the effect of practice, showed only the slightest evidence of learning change. A similar slight reduction in error level occurred over successive 5-sec. means during the trials or observations.

To determine the main sources of variance related to modes of tracking, Duncan (1955) multiple-range tests were computed at the 1% and 5% levels, as shown in Table 1. In this table, two or more means linked by an underline are not significantly different from one another. The table indicates that means for modes were all significantly different from one another at the 5% level, but that the means for eye and head-eye tracking were not significantly different from one another at the 1% level. A similar analysis was made of the means for delays, as given in Table 2. This table indicates that at both the 1% and 5% level, all means were significantly different from one another except those for the 0.2-sec. and 0.4-sec. delay conditions.

Analysis of the significance of the differences between mean tracking error for delay intervals with specific movement modes of tracking, as described earlier by the curves in Figure 6, are given in Table 3. The table, which is based on range tests at the 1% level of significance, indicates again that the effects of delay on the specific modes of eye and head tracking were highly reliable except for the 0.2- and 0.4-sec. delay levels. The reliability of the differences for the combined head and eye tracking was less than that for the specific movement modes.

SUMMARY

This experiment has evaluated the validity of the doctrines of response to distorted vi-

TABLE 3

DUNCAN MULTIPLE-RANGE TESTS AT THE 1% LEVEL OF THE SIGNIFICANCE OF THE DIFFERENCE BETWEEN MEANS FOR DELAYS WITH THE THREE MOVEMENT MODES OF TRACKING

Movement mode	0.0	0.2	0.4	0.6	0.8	1.5
Head	5.3	11.37	10.98	14.89	16.45	29.29
Eye	15.52	15.95	19.38	18.66	22.30	30.54
Head-eye	13.93	18.63	17.53	18.63	20.29	29.56

Note.—Lines, including bracketed dashed lines, underlying common means are not significantly different from one another.

sion relative to their application to specification of the human factors in design of machines and visual measurement and training procedures. Observations confirmed prior work (Smith & Molitor, 1968) in showing that reversal of eye-movement–retinal feedback caused marked rapid unstable vision, to which *S* made little or no adaptation. Such reversal of vision eliminated smooth pursuit movements of the head and eyes and degraded vision to use of saccadic movements. Restriction of vision and visual tracking to reversal of the feedback effects of eye, head, and head-eye motions impaired eye-movement functions more than head movements. Head movements with compensatory visual feedback were also reduced to saccadic movements, but these head saccades were very small and still made possible accurate head tracking.

The experiment specifically tested the hypothesis that the effects of reversal of eye-movement feedback are distinct from reversal of head movement control of visual tracking and localization, as the latter has been studied in inversion and reversal of the visual field by prisms and telescopes mounted in experimental spectacles. The results indicate that reversal of eye-movement–retinal feedback is more severe in causing unstable vision and inaccurate visual localization than anything which has been observed with displaced vision by telescope or prism spectacles mounted on the head. The findings specifically confirm the movement-compensation doctrine of response to displaced feedbacks by showing that such displacement affects head and eye movements differently, and that the capability of *S*s to correct reversal of retinal feedback of eye movements is next to impossible and would require changing the built-in directional specificity of pursuit eye motions in controlling movements of the retinal image.

The present experiment represents a critical test of the theories of displaced vision because it is one of the first to isolate the dynamic feedback of particular response variables related to head and eye motions. This experimental systems procedure was necessary to determine to what extent a given movement can compensate alterations in its own retinal feedback effects. It also was necessary to preclude cross compensatory feedback adjustments between two mechanisms—that is, between the eye and head or between the head and eye. Under such controlled systems conditions, it was found that the capability of the eye system to correct for reversed feedback of ocular movements was next to impossible because the built-in directional specificity of ocular movements in controlling positioning of the retinal image cannot be altered. In contrast, the head movement system could compensate reversals of feedback to some degree because the head does not possess the ingrained directional feedback relations between eye motion and retinal image positioning. The findings indicate that the role of feedback relations between ocular dynamics and retinal stimulation in determining space perception and visual guidance cannot be assessed by tests of effects of optical distortions on head and body movements.

The experiment evaluated the validity of the movement-compensation hypothesis and the visual tactual learning theory in a second way by comparing the effects of delay of retinal feedback on head and eye movements. The classical visual-tactual association hypothesis specifies a role of time factors in guidance behavior and space perception only insofar as temporal contiguity in visual-tactual association can determine learning of space perception. In contrast, the motor-sensory-compensation hypothesis assumes that delay and other feedback timing factors can affect visual guidance just as much as space displacement of retinal feedback. The results confirm this assumption. Support for this view was evidenced also by the fact that retinal feedback delay differed for eye movements and head movements. Relatively, the more accurate head tracking was affected more than the compensatory ocular tracking. Large feedback delays elminated the rapid unstable movements of reversed eye-movement tracking, but nevertheless induced systematic increase in eye-tracking error that varied with the size of the feedback delay interval. The superiority of the head movements was related to the fact that with inverted vision the eye could not maintain local vision, whereas such vision could be achieved when the visual feedback of these movements was reversed. Head-eye tracking, as tested by equating and combining the transduced signals of head and eye motions into a yoked feedback control of a visual target, combined the characteristics of movements observed in the isolated patterns of head and eye tracking.

The results give added meaning to the behavioral feedback interpretation of the critical human factors in defining man-machine and man-systems relationships. According to this interpretation, the effectiveness of machines and devices for human use and the requirements for learning in operating them depend on the degree of compliance in the space, time, and force factors between living motion and the action of machines. Tools and machines introduce an instrumental feedback factor into behavior which occurs as an adjunct to the feedback effects from movement itself and from operational control of the environment which occurs in unaided behavior. These different sources of feedback must correlate with one another in terms of their space, time, and force variables. The greater the discrepancy in feedback between these different sources of feedback, the greater the demands for learning the use of tools and machines and the more limited the effectiveness for adaptation in their operation. Results of the present experiment suggest that the more tools and machines alter the critical retinal feedbacks of the eye itself, the less effective the device for human use and learning.

As indicated by the results on delayed eye-movement–retinal feedback, visual disturbances can result from alterations in feedback timing which are continuously involved in ocular and head control of retinal input. The conclusion may be made that visual perceptual disabilities, such as those found in educational skill, may be based on asynchronism of control of retinal stimulation by eye and head motion. Inasmuch as several levels of dynamic control of retinal stimulation are always involved in vision, many possibilities exist for feedback asynchronism in use of the eyes and in producing feedback disturbances of perception. The systems methods of dynamically isolating response variables in vision for the study of delay effects may represent a solution to the problems of measuring visual functions of the eye in action. The results on delayed feedback indicate that such dynamic methods of visual diagnosis may represent a crucial need in understanding perceptual deficiencies which cannot be reduced to optometric or ophthalmologic defects.

REFERENCES

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrica,* 1955, **11**, 1–42.

FREEDMAN, S. J. *The neuropsychology of spatially-oriented behavior.* Homewood, Ill.: Dorsey Press, 1968.

HAY, J. C. Visual adaptation to an altered correlation between eye movement and head movement. *Science,* 1968, **160**, 429–430.

HAY, J. C., & PICK, H. L., JR. Visual and proprioceptive adaptation to optical displacement of the visual stimulus. *Journal of Experimental Psychology,* 1966, **71**, 150–158.

HELD, R., & HEIN, A. V. Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills,* 1958, **8**, 87–90.

MOWRER, O., RUCH, T. C., & MILLER, N. E. The corneo-retinal potential difference as the basis of the galvanometric method of recording eyemovements. *American Journal of Physiology,* 1936, **114**, 423–428.

ROCK, I., & HARRIS, C. S. Vision and touch. *Scientific American,* 1967, **216** (5), 96–104.

SMITH, K. U., KAPPAUF, W. E., & BOJAR, S. The functions of the visual cortex in optic nystagmus at different velocities of movement of the visual field. *Journal of General Psychology,* 1940, **22**, 341–357.

SMITH, K. U., & MOLITOR, K. Adaptation to reversal of retinal feedback of eyemovements. *Journal of Motor Behavior,* 1968, **1**, 69–87.

SMITH, K. U., & SMITH, W. M. *Perception and motion: An analysis of space-structured behavior.* Philadelphia: Saunders, 1962.

VON HELMHOLTZ, H. L. F. *Handbuch der physiologischen optik.* Hamburg & Leipzig: Voss, 1856–66. 3 vols.

WALLACH, H., KRAVITZ, J. H., & LINDAUER, J. A passive condition for rapid adaptation to displaced visual direction. *American Journal of Psychology,* 1963, **76**, 568–578.

(Received October 8, 1969)

Bureau of Educl. & Psyl. Research
(S. C. E. R. T.)
Date 6.3.81
Acc. J 698

Journal of Applied Psychology
1971, Vol. 55, No. 1, 22–27

EFFECT OF HOURLY OVERPAY AND UNDERPAY INEQUITY WHEN TESTED WITH A NEW INDUCTION PROCEDURE[1]

E. R. VALENZI AND I. R. ANDREWS[2]

Bowling Green State University

This study tested Adams' inequity theory with an induction which did not challenge the workers' job qualifications, eliminating devalued self-esteem as a confounding variable. Thirty-one workers were hired for an hourly pay clerical task in a real job situation and assigned to one of three groups—overpay, underpay, and control—in a before–after design. Contrary to inequity theory predictions and to previous inequity theory experiments, there were no significant work performance differences among the three groups, but 3 of 11 underpay *S*s quit the job while no other *S*s quit. It was concluded that the results of this experiment combined with the results of other experiments strongly suggest that inequity effects previously reported are probably due mainly to the self-esteem variable. It was suggested that research on wage inequity should focus more on variables such as turnover, satisfaction, and recruitment rather than solely on work performance.

Five recent articles (Andrews & Valenzi, 1970; Friedman & Goodman, 1967; Gordon & Lowin, 1965; Lawler, 1968a; Moore, 1969; Pritchard, 1969) have suggested that when perception of overpay inequity is created by challenging the worker's qualifications to perform the job, results are seriously confounded by the simultaneous arousal in the worker of motives to protect self-esteem. The purpose of this study was to test inequity theory (Adams, 1965) with an induction procedure which did not challenge the worker's job qualifications, eliminating devalued self-esteem as a confounding variable.

For a variety of reasons, this study used hourly paid workers rather than piece-rate workers. In his review of wage inequity research, Lawler (1968b) concluded that results from studies that paid workers on a piece-rate basis can be explained by expectancy theory (Vroom, 1964) with only a slight modification based on inequity theory. By using hourly paid workers, the authors avoided this alternate explanation since expectancy theory predicts no difference between hourly paid inequity groups and equity groups.

Second, hourly pay is a more stringent test of inequity theory because of the differential effort required to restore equity. Under the piece-rate condition, perceived overpay inequity can be corrected by limiting quantity of work, thereby reducing outcomes. Even an attempt to improve work quality wouldn't require much effort since the work pace is very slow. With hourly overpay, on the other hand, the worker can influence only the input side of the outcomes/input ratio. Therefore, to correct a perceived inequity, he must increase his work effort rather substantially. In the hourly underpay condition, the worker can decrease his inputs by doing less work and/or poorer quality work. This strategy requires a deliberate intent to perform poorly and has a kind of revenge or get even motive associated with it. It may be psychologically more difficult to rationalize such behavior than it is to correct piece-rate underpay by increased work quantity with some loss in work quality.

Third, no experimental study has been reported that tested inequity theory for the effect of hourly underpayment. Thus, the present study extends previous work on inequity theory in this respect.

Fourth, hourly payment is far more common in the industrial setting then piece-rate payment and therefore is of greater intrinsic interest.

---

[1] This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to I. Robert Andrews who is now at the Department of Economics and Commerce, Simon Fraser University, Burnaby 2, British Columbia, Canada.

The present study is the first experimental test of inequity theory in the hourly overpayment and underpayment condition that: (*a*) created overpay and underpay inequity without complications from job security and self-esteem as confounding variables, (*b*) gave *S*s inequity relative to self and to other salient workers, (*c*) allowed each *S* to come in contact with other *S*s in his work group, and (*d*) manipulated outcomes instead of inputs.

Stated in its most general form, the inequity theory hypothesis states that if Person perceives inequity, the resulting dissonance will motivate him to reduce the inequity. Adams (1965) defines inequity as follows:

> Inequity exists for Person whenever he perceives that the ratio of his outcomes to inputs and the ratio of Other's outcomes to inputs are unequal. This may happen either (a) when he and Other are in a direct exchange relationship or (b) when both are in a direct relationship with a third party and Person compares himself to Other [p. 280].

Among the modes of inequity reduction which Person may engage in, Adams (1965) cites altering inputs, altering outcomes, or a combination of both as possible alternatives. Under the hourly pay system of this experiment the more specific hypotheses tested were: (*a*) If Person perceives himself as overpaid on an hourly basis, then he will increase his work effort leading to better quantity and/or quality of work; (*b*) if Person perceives himself as underpaid on an hourly basis, then he will decrease his work effort leading to poorer quantity and/or quality of work.

METHOD

Subjects

Thirty-one female *S*s between the ages of 18 and 22 yr. who had indicated a preference for clerical work were obtained from the job applicant file of the Financial Aids Office of a large midwestern university. Only individuals who did not have a job and who were not psychology or sociology majors were selected. Care was taken to see that no one lost out on other job opportunities by accepting employment on this project.

Prior to the experimental manipulation, there were no significant dependent variable differences among the three groups to which *S*s were randomly assigned.

Procedure

Each *S* was contacted by telephone and told that her name was obtained from the Financial Aids Office. The *E* explained that he was a research assistant on the survey research project on campus and that he was presently hiring part-time help to prepare data for computer analysis. The *S* was asked if she were still interested in part-time work. If she was interested, she was asked to come to a group meeting with other girls who would be working on the project. In this way, *E* met with six *S*s at a time. The purpose of the group meeting was to make each *S* aware of all the other workers on the project and of the wage they were to receive. At the group meeting, *E* again explained that he was the research assistant on the survey research project and that he was in charge of processing the data. The office was set up to look like a legitimate research operation with piles of questionnaires, IBM cards, and boxes scattered about. At the group meeting, *E* obtained class schedules from each *S*, briefly explained the nature of the project that they would be working on, showed them the nature of the task they were being hired for, and had each *S* complete a "request-to-hire-student-employee" form obtained from the Financial Aids Office. The *S*s were told the job would last at least 6 wk. It was also made clear that the project was financed by a grant from the National Institute of Mental Health (NIMH) and was therefore independent from the Financial Aids Office as far as pay was concerned. This was necessary since the Financial Aids Office determined the pay rate for all student jobs. The *S*s were also told that they would receive their pay in cash upon the completion of each work period and that they would merely sign a receipt for the money they received. This method of payment was justified as a matter of convenience to eliminate a lot of paperwork.

After the hourly rate for the job was given as $1.40 per hr., *E* asked, "Do you feel that this rate is fair for this job?" All *S*s agreed that this was a fair wage for the job. Finally each *S* was scheduled for two, 2-hr. work sessions with 2 or 3 days between the two work sessions.

The task required *S* to transfer X marks from a three-page questionnaire consisting of 72 semantic differential scales to IBM answer sheets. It was explained that this transformation prepared the questionnaires for machine scoring. The pattern of Xs on the questionnaire was determined by using a table of random numbers. Ten different patterns were used so that every tenth questionnaire was identical. Postexperimental interviews indicated that no *S* was aware of scoring the same questionnaire more than once. This task permitted the collection of both quality data (average number of errors per questionnaire) and quantity data (number of questionnaires scored).

The corrected split-half reliability for work quality in Work Session 1 was .50 and for Work Session 2 it was .64. Both estimates were Spearman rank-order correlations and both were significant at the .01 level. To obtain the correlations, the number of errors made on the even-numbered questionnaires was correlated with the number of errors made on the odd-numbered questionnaires within each work session.

It was not possible to obtain the split-half reliability for the work quantity measure since the only data available for quantity was the total number of questionnaires completed in each work session. A test-retest reliability was computed for the quantity measure by correlating the number of questionnaires completed in Session 1 with the number completed in Session 2. The Spearman rank-order correlation was .88, significant at the .01 level.

Unlike the tasks previously used in equity studies, work quality and work quantity were not negatively correlated. Difference scores (Session 2 minus Session 1) for work quality correlated .06 with difference scores for quantity of work, essentially a zero correlation even after a correction for attenuation.

The sensitivity of work quantity on this task as a dependent variable was demonstrated in a pretest study with 28 introductory psychology students. In their first 20-min. work session, students were given a standard experimental set designed to elicit a high level of motivation. In their second 20-min. work session, 8 small cash prizes were offered to those students who could score the most questionnaires without losing accuracy. Roughly speaking, work quantity doubled in the second work session. Moreover, there was no reduction in average work quality.

Design

A before–after design with a control group was used. The *S*s were randomly assigned to one of three conditions (underpay, overpay, or control) by use of a table of random numbers with the restriction that from each group of 6 *S*s that met together, two were assigned to each experimental condition. The *S*s worked individually in an office down the hall from *E*'s.

At the conclusion of the first day's work period, *S* was paid $2.80, required to sign a receipt for the money, and reminded of her next scheduled work period. Upon reporting to the office for her next day's work, an overpay *S* was told the following:

> Good morning (afternoon) Miss Brown.
>
> As you know we are financed by a grant from NIMH. Well, when we applied for the grant we had to submit a budget to account for expenditures of the funds requested. At the time we submitted the budget, we had need for a statistical clerk. As it turned out the new 360 computer does more work than the old 1620 so we don't need the statistical clerk to do some of the computations. We took the money allocated for that function and divided it up among the clerical workers to increase their wage rate. Well, the business office controls the grant money and this is what we got from them (yesterday, day before).

At this point, *S* was shown a memo typed on university stationery from the office of the controller to the project director. The *S* read the memo which said:

> It is felt that eliminating the statistical clerk from the budget and appropriating funds allocated for his function and distributing them among the general clerical help involves the taking of too much liberty with the budget as approved by the office of NIMH.
>
> We advise as close adherence as possible to the original budget as submitted. A change such as that proposed by you should be effected by a formal application to NIMH for budget revision.

After *S* finished reading the memo, *E* continued:

> By the time we submitted the request for a budget change and got approval for it, the project is likely to be done. Besides we are paying you as you work which complicates matters. So what the project director did was to choose a name at random to receive the pay of the statistical clerk so as to conform to the original budget. Now the pay of the statistical clerk is $2.00 per hr. and your name was chosen. *It has nothing to do with you or your work, it's just a matter of chance that you were picked*. It's important that you don't tell anyone else about your pay change. We don't want to cause any bad feelings among other girls working on the project.

After the induction, *E* answered any questions *S* might have and then said, "Now you understand what this means, you will be receiving $2.00 per hr. while everyone else still gets the $1.40 per hr."

The underpay induction was similar to the overpay, except that a messenger was substituted for statistical clerk and the wage of the messenger was $1.20 per hr. The underpay induction finished with the question, "Now you understand what this means, you will be receiving $1.20 per hr. while everyone else still gets the $1.40 per hr." The control *S*s were engaged in routine conversation for a comparable length of time.

It should be noted that this induction procedure was designed to produce wage inequity relative to all other workers in *S*'s work group, relative to self in the first work session and relative to the going rate ($1.40 per hr.) for other student workers holding jobs on campus. It was felt that this manipulation was much stronger than that given Lawler's (1968a) overpay-by-circumstance group since his workers had equity relative to other workers on the same project.

After the workers had completed the second work session, the true nature of the research project was explained in general terms. Experimental *S*s were then asked to tell *S*s on two open-ended questions how they had reacted to the announcement of the wage increase or decrease. In addition, experimental and control *S*s were asked to rate the job on a 7-point wage inequity scale, ranging from a value of 1, very underpaid, to a value of 7, very overpaid. The group mean ratings for underpay, control, and overpay were 2.3, 4.3, and 5.8, respectively. Mean differences between the control group and each of the experimental groups were significant at the .001 level. Additional evidence for the success of the experimental manipulation is presented in the Discussion section.

Before leaving, each *S* was thanked for his participation, was paid $5.00 for the second work session and was asked to refrain from talking about the project with other students.

RESULTS

A quantity difference score was obtained for each *S* by subtracting the number of questionnaires completed in Session 1 from the number completed in Session 2. The difference scores for the overpay group and the underpay group were compared separately against the difference scores for the control group by the Mann-Whitney *U* test (see Table 1). The value of *U* obtained for the overpay versus control comparison was 48; for the underpay versus control comparison, *U* was 47. A *U* of 26 or less would have been significant at the .05 level for each comparison. Since neither comparison was significant, the null hypothesis (that the control group represented the same population distribution on quantity as the overpay and underpay groups) was not rejected.

A comparison of the extreme groups (underpay vs. overpay) produced a *U* value of 45 with a value of 23 or less needed for significance at the .05 level. Thus, not even the extreme groups comparison produced a significant difference.

Quality difference scores were obtained for each *S* by subtracting the quality score for Session 1 from the quality score for Session 2 (see Table 1). The quality difference scores for the overpay and underpay groups were compared separately against the quality difference scores for the control group. Again, neither comparison was significant at the .05 level. The overpay versus control *U* was 45 and the underpay versus control *U* was 41. A *U* value of 26 or less would have been significant at the .05 level.

A comparison of the extreme groups (overpay vs. underpay) on quality yielded a *U* value of 34 with a value of 23 or less required for significance at the .05 level.

TABLE 1

MEAN QUANTITY AND QUALITY DIFFERENCE SCORES FOR THE THREE GROUPS

Treatment group	Quantity	Quality
Overpay	4.40	−.376
Control	4.30	.090
Underpay	5.20	.282

TABLE 2

NUMBER OF PERSONS WHO WERE ABOVE OR BELOW THE MEDIAN ON BOTH WORK QUALITY AND QUANTITY

Treatment group	Above	Below	Total
Overpay	3	0	3
Control	3	3	6
Underpay	2	2	4
Total	8	5	13

The quantity and quality measures were considered simultaneously in the following way. The median quantity difference score and the median quality difference score were found for the entire sample. For each group, the number of *S*s above the median on both dependent measures and the number below the median on both dependent measures was determined (see Table 2). To evaluate the statistical significance of these frequencies, they were entered into three 2 × 2 contingency tables—one for overpay versus control comparison, one for the underpay versus control comparison, one for the underpay versus overpay comparison—and analyzed by the Fisher exact probability test (Siegel, 1956). None of the comparisons were significant: overpay versus control, $p = .24$; underpay versus control, $p = .50$; overpay versus underpay, $p = .35$.

DISCUSSION

Hourly Overpayment

Since these results, no differences between overpaid and equitably paid workers, are contrary to those of previous hourly overpay studies (Adams & Rosenbaum, 1962; Arrowood, as cited in Adams, 1965; Goodman & Friedman, 1958), it is essential to establish that overpay inequity was in fact perceived. Confidence in the success of the experimental induction is based on four pieces of supporting evidence. First, on a postwork questionnaire, workers in the overpaid group rated themselves on a 7-point wage equity scale as overpaid relative to other workers on the project. Second, on an open-ended question which asked, "What were you thinking about immediately after the supervisor told you of the change in pay rate and then left you to start work?" 7 out of 10 workers mentioned overpay inequity as part of their response. Third, when the experimental work situation was described in detail and the critical part of the experimental manipulation was acted out in front of 43 students from an introductory psychology class, 28 of the 43 students responded to the open-ended question described above in terms of wage inequity. (For a fuller explanation of this role projection technique, see Andrews & Valenzi, 1970.) Confirming those results, out of 9 structured paragraphs describing possible reactions of the worker to the induction, a paragraph depicting wage inequity relative to Other was ranked by the 43 student observers as the most plausible reaction to the treatment. Last, one can consider the face validity of the manipulation. Inequity relative to self was fostered by manipulating the wage rate from an equitable level in the first work session to a much higher rate in the second work session. Inequity relative to Other was encouraged in two ways: first each overpaid worker knew that she was making 60¢ per hr. more than the other workers who were familiar to herself on relevant input characteristics (age, education, sex). Each overpaid worker met all the other workers in the group meeting at the time of hire. Second, $2.00 per hr. was considerably more than the standard campus wage ($1.40 per hr.) for that type of work. Taken together, the authors believe that the above data show that the overpaid workers did in fact perceive overpay inequity.

The lack of a performance difference between overpaid and equitably paid workers can be interpreted as evidence for the position that previous overpay inequity studies were confounded by nonequity motivations aroused by the challenge to job qualifications induction (Andrews & Valenzi, 1970; Lawler, 1968a, 1968b; Moore, 1969; Pritchard, 1969). Apparently, when these nonequity motivations are eliminated, the perception of hourly overpay inequity does not lead to performance improvement.

This finding is consistent with Lawler (1968a) who found no performance difference between unchallenged overpaid workers and equitably paid workers while his challenged overpaid group did do significantly more work than the equitably paid group. Lawler concluded that his results, combined with those of Gordon and Lowin (1965), argue strongly that it is the challenge to qualifications, and not the feelings of overpayment, that accounts for the performance differences in the hourly overpay condition.

Hourly Underpayment

This was the first experiment designed to test inequity theory hypoptheses in the hourly underpayment condition. The results did not support inequity theory predictions concerning the effect of underpayment on work performance and, in a sense, were contrary to previous hourly pay studies (overpay) in that perceived inequity did not affect performance.

Since the induction for underpayment was basically the same as the induction for overpayment, evidence for the success of the underpayment induction is essentially the same as the evidence offered above in support of the overpay induction. Additional evidence is the fact that 3 out of 11 underpay *S*s quit the job, 1 immediately upon being informed of the pay change and 2 after working their scheduled work session. No persons in the equitable paid or overpaid groups quit the job. Thus, it is difficult to argue that the lack of an underpayment performance difference was due to a failure of the induction.

It is of interest to note that despite the perception of underpay inequity, only 1 of 11 workers indicated that she decreased effort during the second work session and 2 said that the pay cut made them work harder. Thus, both the work performance data and questionnaire responses indicate that underpayment and the resulting dissatisfaction did not affect performance. This result is consistent with current thinking on the relationship among satisfaction, rewards, and performance (Smith & Cranny, 1968) which states that rewards and satisfaction can influence performance only if mediated by an effort or intention variable.

On the other hand, the fact that 3 of 11 underpay workers quit the job and 5 of the remaining 6 said they thought about quitting, demonstrates that underpayment does have an effect. These turnover data suggest that perhaps attention has been focused on the wrong dependent variables in wage inequity research. The primary effect of underpay inequity may be on turnover and on other forms of leaving the field such as absenteeism and transfer requests to more equitable paying jobs. Similarly, the primary effect of overpay might be on ease of recruitment and the labor pool from which you are able to select workers.

Limitations

The overpay manipulation might have been stronger if the workers thought that their co-workers knew about the pay increase. It is possible that overpayment relative to co-workers is not particularly distressing if the co-workers do not know about it. Co-worker ignorance combined with Person's knowledge that circumstances required one worker to be overpaid could have enabled Person to rationalize the overpayment.

Although underpay *S*s were upset by the pay cut, they may have been reluctant to take it out on the research assistant or the project director since they were merely intermediaries in the pay change. After all, the real culprit was someone in Washington, D. C. This fact was especially apparent in the memo shown to *S*s during the induction. Also underpay *S*s may have been hesitant to restrict production or turn in poor quality work because such behavior has overtones of a revenge or a get even motive associated with it. Middle-class college coeds may have felt that such behavior was inconsistent with their cultural norms of acceptable conduct. In addition, it was obvious how many questionnaires were completed since they were prenumbered consecutively. A purposive restriction in quantity would have been immediately apparent and a fear of reprimand could have caused *S*s to maintain their quantity level.

Generalization of these hourly pay results to the piece-rate condition should be done with caution. To the extent that the challenge to qualifications is a confounding variable in the hourly condition, it also confounds the piece-rate studies that used it. It is possible that performance differences under piece-rate conditions might disappear if self-esteem confounding is eliminated. But it is also possible that in the piece-rate situation, performance differences might still occur because workers are more attentive to the relationship between pay and productivity. Also, as explained earlier in the paper, it is much easier to correct a perceived wage inequity under piece-rate than under the hourly pay condition.

Last, it should be noted that the behavior of full-time permanent employees might be quite different from the behavior of our temporary, part-time employees.

REFERENCES

ADAMS, J. S. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965.

ADAMS, J. S., & ROSENBAUM, W. E. The relationship of worker productivity to cognitive dissonance about wage inequities. *Journal of Applied Psychology*, 1962, **46**, 161–164.

ANDREWS, I. R., & VALENZI, E. R. Overpay inequity or self-image as a worker: A critical examination of an experimental induction procedure. *Organizational Behavior and Human Performance*, 1970, **5**, 266–276.

FRIEDMAN, A., & GOODMAN, P. Wage inequity, self-qualifications, and productivity. *Organizational Behavior and Human Performance*, 1967, 2, 406–417.

GOODMAN, P., & FRIEDMAN, A. An examination of the effect of wage inequity in the hourly condition. *Organizational Behavior and Human Performance*, 1968, **3**, 340–352.

GORDON, B. J., & LOWIN, A. *Qualifications, co-worker characteristics and productivity—An extension and critique of Adams' inequity theory*. (Tech. Rep.) Bell Telephone Laboratories, 1965.

LAWLER, E. E., III. Effects of hourly overpayment on productivity and work quality. *Journal of Personality and Social Psychology*, 1968, **10**, 306–314. (a)

LAWLER, E. E., III. Equity theory as a prediction of productivity and work quality. *Psychological Bulletin*, 1968, **70**, 596–610. (b)

MOORE, L. Effect of wage inequities on work attitudes and performance. Unpublished master's thesis, Wayne State University, 1969.

PRITCHARD, R. D. Equity theory: A review and critique. *Organizational Behavior and Human Performance*, 1969, **4**, 176–211.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

SMITH, P. C., & CRANNY, C. J. Psychology of men at work. *Annual Review of Psychology*, 1968, **19**, 467–496.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

WEICK, K. E., & NESSET, B. Preferences among forms of equity. *Organizational Behavior and Human Performance*, 1968, **3**, 378–416.

(Received October 9, 1969)

rnal of Applied Psychology
1, Vol. 55, No. 1, 28–33

AFFECTIVE RESPONSE TOWARD USES OF RECLAIMED WATER[1]

WILLIAM H. BRUVOLD[2]

University of California at Berkeley

Three Thurstone-type affective attitude scales were developed in order to assess public reaction to high contact uses of reclaimed waste water. The scales were evaluated by field survey research conducted in northern and southern California. Test-retest reliability coefficients ranged from .82 to .79, and scale results met expectations regarding the influence of degree of contact and environment upon attitudes toward waste water reuse. The scales developed should be useful in future applied and theoretical research involving waste water reclamation.

With a rapidly increasing population, there is currently in the United States an urgent need to reclaim domestic and industrial waste water in order to combat further degradation of the environment and to provide sorely needed new sources of water for beneficial uses (Metzler & Russelman, 1968). With recent advances in technology, it is possible to renovate waste water to such a degree that it meets existing domestic water quality standards. However, even though the technical capabilities for thorough waste water renovation are available, there are serious questions regarding public acceptance of reclaimed water for higher contact uses. The idea of consuming water recovered from domestic sewage may be extremely repugnant to many in spite of needs for new sources of water, the imperative to combat degradation of the environment, and current technological expertise.

The controversies created by proposals to fluoridate domestic water (Kirscht & Knutson, 1961) have emphasized the role of public attitudes in water resources matters. Awareness of the importance of public attitudes has led to the initiation of a program of research which will aid water utilities and involved public agencies to deal rationally with these attitudes when planning and initiating new uses of reclaimed water. This research is intended to aid the maximum development of new uses of reclaimed water with a minimum of damaging public controversy. It is believed that these goals can best be attained by anticipatory concern on the part of water utilities and public agencies regarding the attitudes and desires of the consumer population.

One very useful means of obtaining information about public reaction toward specific water reclamation projects and proposals is to assess attitudes toward various uses of reclaimed water. Such data, along with other input regarding public reaction obtained by other procedures, can be of great benefit in helping utilities and agencies to plan proper uses of reclaimed water. The aims of the research reported here were: (*a*) to develop scales assessing the affective component of attitude toward the use of reclaimed water for drinking (primary domestice use), laundry (secondary domestic use), and swimming (primary recreational use); and (*b*) to investigate the reliability and validity of these scales by actual public surveys.

METHOD

Scale Construction

Items. Thurstone (1931) has emphasized the affective component of attitude in defending differential attitude scaling. A contemporary view of how this component fits into overall attitude structure is given by Dillehay (1965). Torgerson (1958) has warned against difficulties involved in attitude measurement using the judgment approach to scale construction and the response method of measurement. Specifically, Torgerson's comments can be interpreted to mean that attitude items scaled for affectivity may be responded to on bases other than affect producing an obfuscated result. Thus, in gathering potential attitude items, every effort was made to include only those that clearly and directly expressed affect toward a specific use of reclaimed water without injecting other attitudinal dimensions such as beliefs about disease producing potential or the lack of need for new water resources.

---

[1] This research was supported in part by Grant W201 from the University of California Water Resources Center, and in part by USDHEW General Research Support Awards 5 S01 FR 5549 and 5 S01 FR 5441.

This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to William H. Bruvold, School of Public Health, Earl Warren Hall, University of California, Berkeley, California 94720.

With these and other well known criteria (Edwards, 1957) as guides, 83, 58, and 82 items were developed dealing specifically with the use of reclaimed water for drinking, swimming, and laundry, respectively.

Subjects. Fifty individuals volunteered to serve as judges for the item scaling procedure. Twenty-seven were males and 23 were females. Eighteen were students, 17 were engineers, and 15 were otherwise employed. All *S*s resided in the San Francisco Bay area. Ages of *S*s ranged from 17 to 57 yr.

Procedure. Items were typed, one to a plain white 3 × 5 in. card, and labeled at random with a one-letter three-digit number combination which appeared in the upper right corner of each card. Order of items within sets, and order of the sets themselves, were randomly arranged before each individual judging session. Actual judgment procedures are described by the instructions read to each *S* by *E*. These instructions are reproduced below.

> Each card in the following group bears a statement describing an attitude about reclaimed sewage water. Consider the statement on each card in turn and decide how favorable or unfavorable it is toward reclaimed water. You are asked to judge the favorableness or unfavorableness expressed by each statement in terms of the following 11 category scale. The distance between integers is equal in the scale shown below.

1	2	3	4	5	6	7	8	9	10	11
extremely unfavorable					neutral					extremely favorable

> Do not respond in terms of your own agreement or disagreement with the statements, but in terms of how favorable you judge each statement to be. Record the number of each statement and your rating of the statement on the appropriate line of the following data sheets. We are not interested in your personal attitude toward reclaimed water but where these statements fall on the continuum described above. The idea is that you are to act as neutral transducer, that is, you are to set your own attitude aside and judge each statement on content only.

After receiving the instructions, each judge rated, for practice only, under the observation of *E*, five statements dealing with the use of reclaimed water for cooking. Questions regarding judgment procedures were fully discussed before *S* proceeded to judge the major item sets. Each *S* worked alone to complete the major judgment task. Integers from 1 to 11 for each item were entered in a specially prepared data booklet. The 11-point scale shown above, and the last two sentences of the instructions were duplicated at the top of each page of the data booklet.

Analysis. Median and *Q* values were obtained for each item using standard methods of analysis associated with the equal appearing intervals procedure (Edwards, 1957). On the basis of these values and other usual considerations (Edwards, 1957), three scales containing 26 items each were constructed. Selected items, and their scale and *Q* values, are shown in Table 1.

Survey Work

Towns. Two towns were chosen for the survey work between which significant differences in public attitudes toward uses of reclaimed water were expected. Town 1, an eastern suburb of San Francisco, had experienced the use of reclaimed water only for golf course irrigation, and this use had been underway for some 2 yr. at the time of the survey. Town 2, an eastern suburb of San Diego, was roughly equal to Town 1 in size, socio-economic level, and type of dwelling unit. However, Town 2 had experienced considerable reuse of reclaimed water for some 8 yr. including golf course irrigation, recreational lakes (boating, fishing), and swimming. Further, Town 2 was located in an arid area having a general scarcity of water, while Town 1 was located in a fertile agricultural valley having no general scarcity of water.

Respondents. Twenty-five respondents were selected from each of the two towns by quota sampling. First, five square blocks were selected in order to be geographically representative of each town. Households were then contacted by direct visit until five individuals from each block agreed to participate in the survey. Not more than three women were selected from any one block. There were 10 male and 15 female respondents in Town 1, and 11 male and 14 female respondents in Town 2. Ages of the respondents ranged from 19 to 61 yr.

Procedure. Interviewing was conducted in the respondent's home at his convenience by the writer and two interviewers working under the auspices of

TABLE 1

SELECTED ITEMS AND THEIR SCALE AND Q VALUES FOR SCALES 1–3

Item no.	Scale value	Q value	Item
			Scale 1
1	10.57	1.41	I am enthusiastic about the idea of drinking reclaimed water.
2	10.27	1.32	I am decidedly pleased with the idea of drinking reclaimed water.
3	9.78	1.60	I would be happy to drink reclaimed water.
4	9.55	1.46	I would be glad to drink reclaimed water.
5	8.65	1.85	Drinking reclaimed water is agreeable to me.
6	8.28	1.88	I would have no objections to drinking reclaimed water.
7	8.04	1.29	Drinking reclaimed water seems like a reasonable idea to me.
8	7.92	1.81	I don't mind drinking reclaimed water.
9	7.70	1.22	I feel that reclaimed water could be used for drinking.
10	7.10	0.71	I am slightly in favor of drinking reclaimed water.
11	6.91	1.25	I suppose it would be okay to drink reclaimed water.
12	6.54	1.79	While I would rather not drink reclaimed water, I think that I might be able to get used to the idea.
13	6.20	1.01	It wouldn't make any difference to me if I were drinking reclaimed water or not.
14	5.95	0.56	I'm neither greatly in favor of nor greatly opposed to drinking reclaimed water.
15	5.13	1.38	I don't know whether I could accept the idea of drinking reclaimed water or not.
16	4.62	1.21	I would be slightly disappointed if I had to drink reclaimed water.
17	4.02	1.35	I'm not very enthusiastic about the idea of drinking reclaimed water.
18	3.76	1.34	I would prefer not to drink reclaimed water.
19	3.62	1.86	I'm not in sympathy with the idea of using reclaimed water for drinking.
20	3.22	1.54	Drinking reclaimed water is a rather disagreeable idea to me.
21	2.83	1.52	Drinking reclaimed water is too radical an idea for me.
22	2.44	1.88	I think drinking reclaimed water is fundamentally unsound.
23	2.00	1.08	It will be a sad day when we have to drink reclaimed water.
24	1.83	1.40	The notion of actually drinking reclaimed water is ridiculous to me.
25	1.39	1.05	The idea of actually drinking reclaimed water must be discarded immediately.
26	1.17	0.71	I would never drink reclaimed water.
			Scale 2
1	10.44	1.18	Using reclaimed water for swimming is an advanced idea which I highly favor
2	10.34	1.24	I am enthusiastic about using reclaimed water for swimming.
3	9.80	1.34	I would be happy to use reclaimed water for swimming.
4	8.96	1.77	I think it is perfectly reasonable to swim in reclaimed water.
5	8.83	1.48	I would have no problems swimming in reclaimed water.
6	8.50	1.52	Swimming in reclaimed water is satisfactory to me.
7	8.22	1.50	I see no reason why one could not swim in reclaimed water.
8	8.19	1.34	Swimming in reclaimed water would be agreeable to me.
9	7.70	1.60	It wouldn't bother me to swim in reclaimed water.
10	7.17	0.97	I am somewhat in favor of using reclaimed water for swimming.
11	6.97	1.18	I really can't criticize the use of reclaimed water for swimming.
12	6.55	1.63	I am not wild about the idea of swimming in reclaimed water, but I could get used to it.
13	6.22	1.08	It wouldn't make any difference to me if I were to swim in reclaimed water.
14	6.04	0.54	I am neither greatly in favor of nor greatly opposed to using reclaimed water for swimming.
15	5.97	0.59	I have no feeling one way or the other about swimming in reclaimed water.
16	4.84	0.74	It might bother me a little to swim in reclaimed water.
17	4.57	1.06	I am slightly bothered by the idea of swimming in reclaimed water.
18	4.31	1.17	The idea of actually swimming in reclaimed water makes me feel a little uncomfortable.
19	3.71	1.71	I would never get enthusiastic about swimming in reclaimed water.
20	3.43	1.66	I wouldn't feel right swimming in reclaimed water.
21	3.17	1.34	Personally, I would have problems swimming in reclaimed water.
22	2.75	1.88	I don't need to swim bad enough to swim in reclaimed water.
23	2.31	1.60	I don't care what they say, I just couldn't get used to swimming in reclaimed water.
24	2.08	1.77	I would certainly hate to swim in reclaimed water.
25	1.59	1.14	I can't think of anything much more unpleasant than swimming in reclaimed water.
26	1.39	1.08	Using reclaimed water for swimming is a terrible idea.

Table 1—(*Continued*)

Item no.	Scale value	*Q* value	Item
			Scale 3
1	10.63	1.22	Using reclaimed water for laundry is an advanced idea which I highly favor.
2	10.33	1.29	I am enthusiastic about the idea of using reclaimed water for laundry.
3	10.16	1.23	I am definitely pleased with the idea of using reclaimed water for laundry.
4	9.70	1.76	I would be very satisfied to use reclaimed water for laundry.
5	9.09	1.21	It would be a good idea to use reclaimed water for laundry.
6	8.50	1.36	Using reclaimed water for laundry is agreeable to me.
7	8.21	1.23	Using reclaimed water for the laundry seems like a reasonable idea to me.
8	8.00	1.09	I feel that reclaimed water could be used for laundry.
9	7.71	1.20	I wouldn't mind using reclaimed water for laundry.
10	7.27	0.99	I am somewhat in favor of the idea of using reclaimed water for laundry.
11	6.70	1.82	Maybe I could get used to reclaimed water for laundry.
12	6.65	1.91	While I am not wild about the idea of reclaimed water in general, maybe I could use it for laundry.
13	6.01	0.53	The idea of using reclaimed water for the laundry doesn't affect me one way or the other.
14	5.98	0.53	I'm neither greatly in favor of nor greatly opposed to using reclaimed water for laundry.
15	5.50	1.28	I don't know whether I could accept the idea of using reclaimed water for laundry or not.
16	4.87	0.67	The idea of using reclaimed water for the laundry is a little more unfavorable than favorable to me.
17	4.64	1.19	I would be slightly disappointed if I had to use reclaimed water for laundry.
18	4.14	1.74	I would feel somewhat concerned if I had to use reclaimed water for the laundry.
19	3.79	1.54	I would prefer not to use reclaimed water for laundry.
20	3.43	1.44	I have reasons why I would rather not use reclaimed water for the laundry.
21	2.40	1.40	I have plenty of objections to the use of reclaimed water for laundry.
22	2.38	1.66	I don't think reclaimed water would even be good enough for the laundry.
23	2.08	1.66	I will be most unhappy if I have to use reclaimed water for my laundry.
24	1.81	0.96	Using reclaimed water for laundry is very disagreeable to me.
25	1.65	1.20	As far as I'm concerned, reclaimed water should definitely not be used for the laundry.
26	1.12	0.62	Reclaimed water for laundry disgusts me.

the California State Department of Public Health. The first phase of the survey was completed late in May 1969. The interview form employed in this phase of the work consisted of a series of questions regarding water resources matters and the three scales described herein. Scales were administered after careful instruction to read all items first, and then to mark only those items closest to the respondent's personal attitude. The same random item arrangement was used for each scale. The order of scale administration was randomized separately for each respondent.

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR THE FIELD SURVEY DATA

Data	Replication 1			Replication 2		
	Scale 1	Scale 2	Scale 3	Scale 1	Scale 2	Scale 3
Town 1						
$\bar{X}$	4.95	5.88	6.32	5.40	6.53	7.04
SD	2.20	2.45	2.37	2.42	2.37	2.10
Town 2						
$\bar{X}$	5.96	7.31	7.51	6.24	7.37	7.87
SD	2.05	1.52	1.77	2.24	1.82	1.68

The second phase of the survey was conducted late in July 1969. The 50 respondents were mailed letters preceding the follow-up study thanking them for their earlier participation and requesting their cooperation in the second and final phase of the research. All 50 participated in the follow-up study.

TABLE 3

TEST-RETEST RELIABILITY COEFFICIENTS FOR THE THREE SCALES

Scale	Coefficients
1	.82
2	.80
3	.79

TABLE 4

ANALYSIS OF VARIANCE FOR THE FIELD SURVEY DATA

Source	*SS*	*df*	*MS*	*F*
Towns				
1	77.00	1	77.00	3.65*
SwG	1013.14	48	21.11	
Replications				
2	12.79	1	12.79	6.77**
12	2.33	1	2.33	1.23
2 × SwG	90.54	48	1.89	
Scales				
3	130.64	2	65.32	35.12***
13	.62	2	.31	—
3 × SwG	178.43	96	1.86	
23	.39	2	.20	—
123	.59	2	.30	—
23 × SwG	48.93	96	.51	

Note.—SwG = *S*s within groups.
* $p < .075$.
** $p < .025$.
*** $p < .001$.

Interviewing was performed in the respondent's home by the two State Health interviewers. A different random order of item arrangement was employed in the follow-up study, with order of scale administration again randomized separately for each respondent. A few brief questions regarding water resources matters preceded scale administration.

RESULTS

The respondent's score for each scale administration was taken as the median of all checked items. Means and standard deviations for each scale, replication, and town are shown in Table 2. Table 3 records the test-retest reliability coefficients for each scale. The coefficients of Table 3 are each based upon 50 pairs of scores.

Statistical analysis of differences between the means of Table 2 posed two problems. The first involved the many possible pairwise comparisons and the second the legitimacy of comparisons between different scales. The first problem is well handled by an analysis of variance approach. Also, since the three scales were constructed by the same judgment procedure on a common quantitative base giving highly similar, if not identical, meaning to response scores on the affective continuum as here optionally defined, the decision was made to use analysis of variance as a defensible expedient to efficiently assess differences between all the means of Table 2.

The selection of two towns, the use of three scales of common quantitative base, and the two replications made possible a three-factor analysis with repeated measures on the last two factors listed (Winer, 1962). The results of this analysis are shown in Table 4.

DISCUSSION

The reliability coefficients of Table 3 show that individual scores obtained from the two administrations were reasonably stable. The magnitude of these coefficients would have been higher were it not for a few respondents from both towns whose attitudes had become markedly more positive by the second administration. These changes were apparently true changes in attitude indicating that the degree of instability revealed by the coefficients of Table 3 was not wholly the result of error variance. Comments made to the interviewers indicated that some of these changes were due to discussions with relatives and friends stimulated by the first interview session, and some were apparently due to an increased faith in American technology due to the successful moon landing of July 1969 ("If we can land men on the moon we sure ought to be able to purify sewage water.") The shift toward more positive attitudes is clearly revealed in the means of Table 2 and the variance analysis of Table 4.

Validity is a more vexing problem than reliability for attitude measurement. The lack of criteria against which the results of affective attitude scales can be validated seems to make imperative adoption of some form of a construct validity approach. Such was the case with the present research. It was expected on the basis of initial reasoning that Town 1 respondents should have less favorable attitudes toward the three uses of reclaimed water than Town 2 respondents, and that favorability of attitudes toward reuse of water should be inversely related to degree of contact. The latter expectation means that attitude scores for the present research should be least favorable for Scale 1, intermediate for Scale 2, and most favorable for Scale 3. Support for both expectations was provided by the present study as evidenced by results of the analysis of variance and the means of Table 2. The fact that the F ratio for towns did not quite meet the customary level required for significance is ameliorated by the sample size and by the consistency of expected and observed mean differences between towns.

Thus, the results of this study have shown that the scales developed can perform adequately in field research with the general public. Therefore, the scales should be useful both in research designed to aid planning for new uses of reclaimed water and in theoretically oriented research designed to assess relations of the affective component of attitudes with beliefs, knowledge, and behavior regarding use of reclaimed water.

REFERENCES

DILLEHAY, R. C. Attitudes and beliefs. In A. L. Knutson (Ed.), *The individual, society, and health behavior*. New York: Russell Sage, 1965.

EDWARDS, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957.

KIRSCHT, J. P., & KNUTSON, A. L. Science and fluoridation: An attitude study. *Journal of Social Issues*, 1961, **17**, 37–44.

METZLER, D. F., & RUSSELMANN, H. B. Wastewater reclamation as a water resource. *Journal American Water Works Association*, 1968, **60**, 95–102.

THURSTONE, L. L. The measurement of social attitudes. *Journal of Abnormal and Social Psychology*, 1931, **26**, 249–269.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received October 20, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 34–36

STABILITY OF THE STRONG VOCATIONAL INTEREST BLANK FOR MEN[1]

CHARLES B. JOHANSSON[2] AND DAVID P. CAMPBELL
Macalester College *University of Minnesota*

Inspection of test-retest reliability studies for the Strong Vocational Interest Blank (SVIB) revealed that two factors were imporant when considering the stability of this instrument: *S*s' initial test ages and length of test-retest intervals. To further investigate these two factors, median test-retest correlation coefficients were computed for over 1,400 college educated males who had been tested with the SVIB at various ages and retested at varying intervals. Greater stability resulted the older the *S* at initial testing and the shorter the time lapse to retesting. Stability decreased as the test-retest interval lengthened or the initial test age of *S* decreased.

In the field of interest measurement, numerous studies have been done attesting to the high test-retest reliability of the Strong Vocational Interest Blank (SVIB). Inspection of the reliability coefficients presented by Campbell (1966) indicates that two factors are important when considering the stability of the SVIB—age at first testing, and the length of time between test and retest.

This study looks more closely at the influence of these two factors on the test-retest stability of the SVIB.

Subjects

Item responses for the SVIB were available for six male groups. Each group was initially tested just prior to, during, or shortly after attending college, and retested at various intervals ranging from 2 wk. to 23 yr.

A 2-wk. test-retest sample ($N = 140$) of students from an introductory psychology class at the University of Minnesota consisted mostly of sophomores, ages 19, 20, or 21 yr., from the College of Liberal Arts.

A 3-yr. test-retest group ($N = 334$) consisted of students attending Harvard University who were tested late in their freshmen year (1960 or 1961) and retested as graduating seniors. These *S*s were mostly 17, 18, or 19 yr. old at initial testing. They form part of a longitudinal study currently being conducted by Stanley King and Bruce Finnie at Harvard.

An 8-yr. test-retest sample ($N = 171$) were graduates of the University of Minnesota in several professional areas. They took the test initially as high school seniors (early 1950s) and all were retested 4 yr. after college by Vera Schletzer (1963). Initial test ages were 17 and 18 yr. and retest intervals were from 8 to 10 yr.

The remaining three groups were tested and retested by E. K. Strong (1955). One group consisted of students tested while in graduate school at Stanford University in 1927 and retested in 1932 and again in 1949. Initial test ages were mostly 21, 22, and 23 yr. Test distribution was as follows:

Taking all three tests	$N = 162$
Taking only two tests	$N = 94$
Total	$N = 256$

The second sample tested by Strong from Stanford University were college seniors tested in 1927 and retested three times: 1932, 1939, and 1949. Test distributions were as follows:

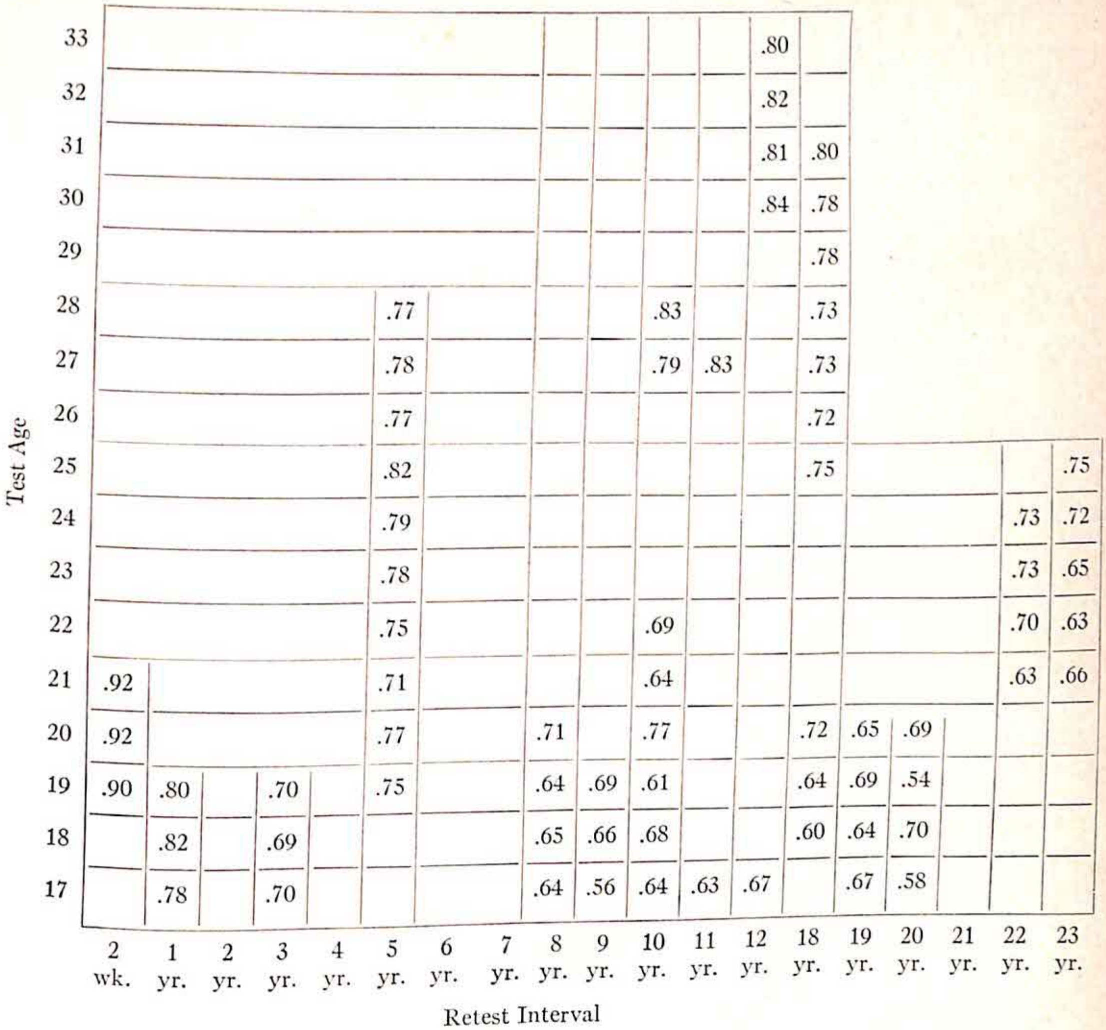
Taking all four tests	$N = 143$
Taking three tests	$N = 82$
Taking only two tests	$N = 37$
Total	$N = 262$

The last group tested by Strong consisted of Stanford University students tested as fresh-

---

[1] This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Charles Johansson, 301 Old Main, Macalester College, St. Paul, Minnesota 55101.



Test Age	2 wk.	1 yr.	2 yr.	3 yr.	4 yr.	5 yr.	6 yr.	7 yr.	8 yr.	9 yr.	10 yr.	11 yr.	12 yr.	18 yr.	19 yr.	20 yr.	21 yr.	22 yr.	23 yr.
33													.80						
32													.82						
31													.81	.80					
30													.84	.78					
29														.78					
28						.77					.83			.73					
27						.78					.79	.83		.73					
26						.77								.72					
25						.82								.75					.75
24						.79												.73	.72
23						.78												.73	.65
22						.75					.69							.70	.63
21	.92					.71					.64							.63	.66
20	.92					.77			.71		.77			.72	.65	.69			
19	.90	.80		.70		.75			.64	.69	.61			.64	.69	.54			
18		.82		.69					.65	.66	.68			.60	.64	.70			
17		.78		.70					.64	.56	.64	.63	.67		.67	.58			

Retest Interval

FIG. 1. Median test-retest correlation.

men in 1930 and retested in 1931, 1939, and 1949. Test distributions were as follows:

Taking all four tests	$N = 124$
Taking three tests	$N = 92$
Taking only two tests	$N = 67$
Total	$N = 283$

The above six groups contained over 1,400 *S*s and almost 3,800 tests. All tests were scored on male scales for the SVIB (Campbell, 1966). Pearson product-moment correlation coefficients were computed between the test and retest for each of the 59 SVIB scales. Next, all test-retest correlation coefficients were analyzed in terms of a bivariate distribution of test age versus retest interval. Figure 1 represents the median correlation, based on sample sizes greater than 20, as a function of test age and retest interval.

For very short intervals (2 wk.) the test-retest reliability of the SVIB is about the same as the split-half reliability (*r*s in the low .90s). Test-retest intervals of 1–5 yr. indicate reliability coefficients in the high .70s for most test ages. Long intervals, 10 yr. and greater, show that the initial test age was the more important; the following coefficients would be expected for long test-retest intervals: .65 if tested during the late teens, .70 if tested in the early twenties, .75 if tested in the late twenties, and .80 if tested after age 30.

Table 1 collapses the initial test ages and the test-retest intervals into a more readable

TABLE 1

MEDIAN TEST-RETEST CORRELATIONS

Test age	Test-retest interval				
	2 wk.	1–5 yr.	6–10 yr.	11–19 yr.	20 & over
16–19	.90	.75	.64	.64	.58
20–22	.92	.75	.70	.68	.67
23–25	—	.79	—	.75	.73
26 & over	—	.78	.81	.80	—

form. The trend was quite clear that when the test-retest interval is held constant, the median reliability is higher in direct relation to the age of *S*, that is, the median test-retest correlations increase as one reads down the columns. Also, for a given age of *S*, the median reliability is lower the longer the interval between test and retest, that is, the correlations decrease as one works across the table.

REFERENCES

CAMPBELL, D. P. *Manual for the Strong Vocational Interest Blank for Men and Women*. Stanford: Stanford University Press, 1966.

SCHLETZER, V. M. A study of the predictive effectiveness of the Strong Vocational Interest Blank for job satisfaction. Unpublished doctoral dissertation, University of Minnesota, 1963.

STRONG, E. K., JR. *Vocational interest 18 years after college*. Minneapolis: University of Minnesota Press, 1955.

(Received October 22, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 37–41

READABILITY OF TECHNICAL TRAINING MATERIALS PRESENTED ON MICROFICHE VERSUS OFFSET COPY[1]

THOMAS S. BALDWIN AND LARRY J. BAILEY[2]

Department of Vocational and Technical Education, University of Illinois

The purpose of the study was to develop psychometric instruments of known reliability and content validity for measuring an individual's ability to process each of several types of information through the visual modality, and to conduct an experimental study in which the instruments developed were used as the dependent variable. Twelve reading tests were constructed using each of four types of technical training materials. Three forms of each test master were reproduced: offset copy, positive image microfiche, and negative image microfiche. A control group received the offset copy form, and positive and negative image forms were administered to the two experimental groups. An analysis of variance revealed statistically significant differences in favor of offset copy for 3 of the 12 subtests.

The existing widespread use of microfilm and the probable growth of this method of information storage are well recognized. Major reasons for the wide acceptance of this medium are the practical advantages of storage, retrieval, and economy. A very real, but not readily apparent, problem has to do with the influence that the medium itself has on the information processing ability of individuals who use microfilm as a major reference material.

A comprehensive review of literature related to research applications of microfilm has been completed by Field (1968). Following the review, the author concluded that "very little recent, definitive work has been published [p. 163]." No attempts have been reported in the literature to compare positive and negative microfiche with an offset copy form of the same reading materials. A possible explanation for the lack of such studies is the necessity for conducting well-designed experimental investigations which employ adequate psychometric techniques.

The purposes of this study were to: (*a*) develop psychometric instruments of known reliability and content validity for measuring an individual's ability to process each of several types of information through the visual modality, and (*b*) conduct an experimental study in which the instruments developed were used as the dependent variable. The study was designed to determine the influence of each of three methods of presentation (independent variable) on information processing ability.

The null hypothesis states that ability of readers to assimilate and utilize information contained in technical training materials is the same for each of the following methods of presentation: (*a*) material presented as black on white offset copy, (*b*) material presented via microfiche with positive image, and (*c*) material presented via microfiche with negative image. From the null hypothesis, two research hypotheses are derived. First, reader performance using black and white offset copy will be significantly better than performance on either positive or negative microfiche. The second research hypothesis is nondirectional and states that reader performance will not be significantly different for positive versus negative microfiche.

METHOD

In order to test the hypothesis of no significant differences in the ability of readers to assimilate and utilize information contained in technical training materials presented by three different methods, it was necessary to develop psychometric instruments which were representative of Air Force training materials.

---

[1] The research reported in this paper was conducted by the Department of Vocational and Technical Education, University of Illinois under Contract No. F33615-68-C-1522 with the Behavioral Sciences Laboratory, Wright-Patterson Air Force Base, Ohio. An expanded version of the study reported herein was submitted by the authors as a final project report in fulfillment of the requirements of the research contract.

This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Larry J. Bailey, who is now Assistant Professor, Technical and Industrial Education, Southern Illinois University, Carbondale, Illinois 62901.

Development of Psychometric Instruments

Psychometric instruments were developed to evaluate *S*s' ability to perform reading tasks utilizing each of four types of technical training materials: (*a*) materials requiring the reading of continuous prose and short paragraphs; (*b*) material requiring the identification of and discrimination among precise symbols and figures; (*c*) material utilizing the skills employed in reading charts, tables, graphs, maps, and isometric and orthographic drawings; and (*d*) material requiring the skills utilized in reading schematics, block diagrams, wiring diagrams, flow charts, and nomenclature verification. Following is an elaboration of the four categories with a description of the representative tasks for each type of material.[3]

Type 1

Narrative-continuous prose (*Tests 1 and 2*). Material for this section was selected from Air Force student study guides concerning the subject of survival training. The content was rewritten as continuous, uninterrupted prose by eliminating figures, headings, questions, etc. After the text had been edited and typed, the manuscript was modified by deleting three words on each page and replacing the words with blanks. Deleted words were selected subjectively on the basis of their relevance to context and assumed ease of identification. The task is to read the material and supply the missing words (Part 1–50 min., Part 2–25 min.).

Narrative-short paragraph (*Test 3*). The same content material was used for this part as was used for the continuous prose exercises. Two and three sentence paragraphs were isolated from the text. The short paragraphs were then modified by changing a word in each paragraph. This technique is similar to that used for the Chapman-Cook Speed of Reading Test (Chapman, 1924). The *S*s are to read the paragraph and identify the word which is inconsistent with the meaning of the paragraph (25 min.).

Type 2

Figure identification (*Test 4*). For each item, *S* checks which of five numbered geometrical figures or pictures in a row is identical to the given figure at the left end of the row (5 min.).

Length estimation (*Test 5*). Each item consists of two parts. Three curved or angular lines are drawn between these two points. The task of the examinee is to select the shortest of these lines (3 min.).

Symbol translation (*Test 6*). This is a test of *S*s' ability to translate symbols into alphabetic characters. A legend illustrates alphabetic letters and their corresponding symbols (typewriter characters). The test items consist of short series of symbols. The task is to identify the alphabetic letter that corresponds to each of the given symbols (10 min.).

Type 3

Graphs (*Test 7*). Speed and accuracy in reading a graph is measured by this test. Minutes are marked on the horizontal scale of the graph and temperatures are shown on the vertical scale. Four curves plot the relationship of temperature to time under certain experimental conditions. For each item, the minutes are given. The task is to find the temperature corresponding to the given number of minutes for each of the four curves (7 min.).

Tables (*Test 8*). This is a measure of the ability to locate and record values found on a table. A table is presented which contains drill bit sizes and their corresponding decimal equivalents. The task is to refer to the table and furnish the decimal equivalent for each of the given drill sizes (4 min.).

Mechanical drawing (*Test 9*). A pictorial drawing (three dimensional) of an object is presented along with three separate views of the object. Each surface of the pictorial is represented by an alphabetic letter. Lines and surfaces on the two dimensional views are represented by numbers. For each lettered surface in the pictorial view, *S* is to find the corresponding numbered line or surface in each of the three separate views (5 min.).

Type 4

Schematics (*Test 10*). Each item consists of a network of lines as in an electrical-circuit diagram having many intersecting and intermeshing wires with several sets of terminals. The task is to trace the lines and to determine for which pair of terminals, marked S (start) and F (finish), there is a complete circuit through a circle at the top (15 min.).

Flow diagrams (*Test 11*). The examinee sees diagrammatic sections representing city maps. The streets are blocked at various points by barriers represented by circles. The examinee must plan routes between given points in such a way that no roadblocks need to be crossed. The task is to follow a path and find the shortest route as quickly as possible (7 min).

Number verification (*Test 12*). The *S* inspects pairs of multidigit numbers and indicates whether the two numbers in each pair are the same or different (6 min.).

Materials selected for inclusion in these instruments were largely derived from Air Force training matter. Content from other sources was determined to have the necessary face validity to represent Air Force material. Preliminary forms of all instruments were administered to a sample of students drawn from the same population as those used for the final testing. This pretesting provided data from which to construct the final forms of the test instruments. Time limits for each test were established that were at least two standard deviations above the mean time for completion, as computed from the pretest data.

---

[3] Several of the test instruments utilized for this research were reproduced or adapted from the Kit of Reference Tests for Cognitive Factors, copyrighted by Educational Testing Service, Princeton, New Jersey. The reference kit was developed under Office of Naval Research Contract Nonr-2214 (00), Project Designation NR 151-174. Reproduction of these test instruments is permitted by and for United States Government use.

Test Forms and Equipment

A master of each test was developed from which the final forms were reproduced. Following completion of the master tests, Itek plates were produced and the black on white offset copy was printed. The same master tests were then used to film the positive and negative image microfiche forms according to military specifications. Microfiche readers used for the study were selected by the contract monitor and were deemed to be representative of the readers in use throughout the Air Force.

Reliability Procedure

The final forms of all test instruments were characterized as primarily speed tests. Therefore, the reliability of each test was determined by a test-retest procedure. Half of each test was administered on successive days to a sample of 42 *Ss*. The offset copy form, only, was used for the reliability testing. All tests were scored by counting the correct number of responses for each exercise. After reliabilities were computed, the Spearman-Brown prophecy formula was used to estimate the reliabilities for the total test length.

Table 1 presents reliabilities for the 12 testing instruments. It can be seen that reliabilities are generally quite high with the exception of Test 10, flow diagrams, which has a reliability of only .59. It would be necessary to increase this test to 2.78 times its present length to achieve a reliability of .80. Reliabilities for the remaining 11 tests range from .76 to .93, all acceptable for use in studying comparisons between groups of individuals.

Selection of Experimental Subjects

The sample for the experimental study was selected from students in Personnel Awaiting Training Status (PATS) at Chanute Air Force Base, Illinois. The total sample was drawn from the population of PATS by a quasi-random procedure. Each Commander of 13 student squadrons was asked to randomly select 8 students per day for testing throughout the duration of the study. The *Ss* were then randomly assigned to one of three sample groups. A total of 133 *Ss* were tested according to the following assignments: 43–positive image sample; 45–negative image sample; and 45–offset copy sample. To test the hypothesis that samples were equivalent with respect to general intellectual ability, an analysis of variance of Armed Forces Qualification Test (AFQT) scores was computed. The AFQT was designed to evaluate enlistees or draftees prior to their entering the Armed Forces and is a composite index of general learning ability. Results of the analysis of variance showed that there were no significant differences among the three groups in terms of AFQT scores ($F = .692$, $df = 2/130$, $p > .05$).

Test Procedures

The total test battery required approximately 4 hr. to administer. Project personnel administered all

TABLE 1

TEST MEANS AND STANDARD DEVIATIONS

Subtest	Offset copy			Positive		Negative	
	$\bar{X}$[a]	*SD*	r_{xx}[b]	$\bar{X}$[c]	*SD*	$\bar{X}$[d]	*SD*
1. Narrative (50 min.)	37	15.2	.92	38	19.1	40	20.7
2. Narrative (25 min.)	19	7.5	.85	19	8.7	20	11.2
3. Narrative—short paragraph	27	11.8	.79	27	11.7	30	13.3
4. Figure identification	61	12.5	.90	54	9.0	52	8.7
5. Length estimation	30	6.0	.90	28	8.8	27	6.2
6. Symbol translation	141	28.4	.79	114	28.3	117	30.7
7. Graphs	34	10.8	.93	28	9.6	29	9.9
8. Tables	17	5.7	.76	16	5.7	15	6.0
9. Mechanical drawing	27	16.4	.90	33	18.3	36	20.4
10. Flow diagrams	19	5.5	.59	18	4.1	20	5.1
11. Schematics	10	5.7	.88	11	5.7	12	6.4
12. Number verification	30	6.9	.91	29	6.9	29	7.3

[a] $N = 45$.
[b] Test-retest, corrected to full length using Spearman-Brown prophecy formula. $N = 42$.
[c] $N = 43$.
[d] $N = 45$.

instruments using standardized test procedures and instructions. To counterbalance the effect of fatigue, the instruments were administered in three different sequences within each group. In addition, *Ss* were permitted to take a 10-min. break at the end of the first hour, 20 min. at the end of the second hour, and an additional 10 min. following the third hour of testing.

RESULTS

To determine the accuracy of the a priori groupings of the tests into the four categories established at the beginning of the study, the intercorrelations among the 12 tests were submitted to a principal axis factor analysis and extracted factors were rotated using a Varimax rotation. By comparing the communalities obtained from the factor analysis to the reliabilities for each of the 12 tests, it was evident that a number of tests in the battery contained considerable specific variance. For these reasons, a separate analysis of variance was computed for each test rather than for test scores combined according to either the a priori categories or the empirical categories determined by the factor structure.

Means and standard deviations for each group on each test are presented in Table 1. The null hypothesis that the ability of readers to assimilate and utilize information contained in technical training materials is the same for each method of presentation was accepted for 9 of the 12 tests. The null hypothesis was rejected at the .05 level of significance for Test 7, graphs ($F = 3.5$, $df = 2/130$); at the .01 level for Test 4, figure identification ($F = 10.48$, $df = 2/130$); and at the .01 level for Test 6, symbol translation ($F = 11.29$, $df = 2/130$). For these three tests the research hypothesis, which states that the method of presentation does affect the readability of the material, is accepted.

Since the null hypothesis was rejected for Tests 4, 6, and 7 by the analysis of variance, further analyses were conducted to determine the significance of differences between pairs of means. A t test was computed for each of these three tests to determine if the mean score for offset copy was significantly higher than for the closest mean for microfiche, whether positive or negative. Also, tests were computed to determine if the mean score for positive microfiche was significantly different than for negative microfiche. One-tailed tests of significance were used in testing differences between offset copy and microfiche, since it was hypothesized that any advantage in readability should favor offset copy. While there was some reason to believe that positive image microfiche would be more readable than negative image microfiche, the evidence was not considered to be sufficient to warrant a directional hypothesis. Hence, a two-tailed test of significance was used.

The t tests for the significance of the differences between the means presented in Table 1 indicate that the mean score for offset copy is significantly larger than the mean for either positive or negative microfiche for Test 4, figure identification; Test 6, symbol translation; and Test 7, graphs. In each case, the difference was significant at the .02 level. None of the differences between means for positive and negative microfiche were significant.

SUMMARY AND CONCLUSIONS

The primary purpose of this research was to determine the influence of information presented on microfilm, on the ability of Air Force *Ss* to read such materials. Comparisons were made between information presented on microfiche and information presented on offset copy, the method most widely used for current Air Force training materials. A search of existing Air Force training materials was conducted and four types of materials were identified. Twelve tests were developed to measure these four types of training materials, 3 tests for each type of material.

The 12 tests were administered to three different groups of Air Force *Ss*. One group received the material presented on offset copy, the second group received the identical material on positive image microfiche, and the third group received the material on negative image microfiche.

Statistical analysis of the data revealed that in 3 of the 12 tests the mean score for *Ss* receiving material via offset copy was significantly greater than the mean score for *Ss* receiving the material via either positive image or negative image microfiche. For the remaining 9 tests there were no significant differences among the three groups. In none of the

12 tests were mean scores for positive image microfiche significantly different from mean scores for negative image microfiche.

Implications for Further Research

This study dealt with one problem bearing on the feasability of using microfilmed materials for instructional purposes. Further research should be directed toward two problem areas: (*a*) investigations of methods for increasing the reading efficiency of microfiche materials and equipment, and (*b*) investigating applications of microfiche materials for education and training purposes. Specifically, further research is needed to determine the influence of the following variables on efficiency of reading: image sharpness, ambient lighting, screen image illumination, surround pattern of image, screen glare, screen angle, color of image, type size, and type form. Also, comparative studies should be made among different types of microfiche readers to identify hardware characteristics that contribute to reading efficiency. Research should also be directed toward determining the relationship of *S*s' preference for film polarity to reading efficiency and reader satisfaction. An informal observation during the experimental testing suggested that *S*s reading the microfiche test forms were more highly motivated than *S*s reading the offset copy. This suggests a study covering a significantly greater length of time, during which *S* is required to use microfiche, to investigate the level and stability of motivation. Finally, research is needed to compare the learning rates of students using microfilmed materials with students using materials reproduced on offset copy.

REFERENCES

CHAPMAN, J. C. *Chapman-Cook Speed of Reading Test*. Educational Test Bureau, 1924.

FIELD, J. V. Human engineering aspects of the design of a microfilm reader. In G. A. Bloch, G. G. Gentry, & J. V. Field (Eds.), *Two studies of the effect of film polarity on patent examiners' performance.* Washington, D. C.: National Bureau of Standards, 1968.

(Received October 30, 1969)

urnal of Applied Psychology
71, Vol. 55, No. 1, 42–45

SIMULATION STUDY OF COMPETITION IN AN "OPEN WORLD"[1]

JOHN L. KENNEDY[2]

Department of Management, Graduate School of Business Administration
University of Southern California

Five three-man teams competed in a business game in an environment in which behavioral access to the simultaneous behavior of a competing team by means of closed-circuit TV and sound could be purchased. Switching to other teams made possible nearly continuous surveillance of all the competitors. Preliminary results describe the various adaptations of the teams to this unusual technological capability. The article also contains speculations on the implications of such research for the worlds of business and international relations.

This article describes a preliminary study or exploration of the consequences of making available low-cost "behavioral access" as a regular condition of competition. Behavioral access, in this case, consisted of a closed-circuit TV picture and associated sound of the simultaneous behavior of a selected competitor. Only one other team could be observed at a given time, but the observing team could, for a fee, switch from one to another team at any time. The competing entities were five three-man teams of undergraduate students in a course entitled "Human Problems in Industry," the 3-hr. laboratory period of which was devoted to team competition in a "business game," the Research and Development Game. Twelve weekly 3-hr. sessions were conducted during the fall term of 1962–63. At the end of the first six sessions, a winner (the team that made the most fictitious money) was declared and the game was played a second time starting with the initial conditions of financial equality.

Several important concepts have been mentioned and need elaboration at this point. The first is behavioral access. We wish to differentiate this concept from the concept of inspection. Behavioral access means access to the actual process of making plans, reaching decisions, and taking actions. It is obtained by unobtrusively getting inside the on-going behavior of a group or team, in this case by fixed-installation television camera and microphone. Inspection, on the other hand, is usually confined to the notion of counting identifiable products, such as number of missile bases, airfields, aircraft by type, etc. The inference of intent from inspection of products is a risky business—the same product may be planned for and utilized for a variety of purposes. Access to the plans and strategies resulting from the differing ways information is processed may be a more powerful tool for estimating intent, and, when used in a mutual or interacting fashion, may create a new form of competition, which is based upon comparative productivity rather than secrecy. Behavioral access is a tool for mutual education. It sets a lower limit on competition. It represents a minimum amount of cooperation to establish an arena of international, cultural, or business competition which might exhibit self-corrective properties.

The second important concept is that of the laboratory simulation study. We shall mean by simulation the process of studying the behavior of people in a real-world environment by creating an operating model of the real-world reference environment in the laboratory. This may be done with or without computers. The computer provides a unique and necessary capability for high speed, high volume information processing sometimes at a lower cost than a noncomputerized operating model and thus computers are often associated with simulation efforts. The study reported here did not require a computer.

---

[1] This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to the author, Department of Management, University of Southern California, University Park, Los Angeles, California 90007.

The third concept is that of the team. By team, we shall mean a group of people who expect to work together toward a common goal for an extended period of time. We view our three-man teams as miniature organizations. They were self-organizing in the sense that, although *E*s assembled the teams on the basis of certain criteria, the team members themselves determined how they should organize to cope with the environment.

Laboratory Environment

The study reported here is part of a continuing program of research on cooperation and competition with simulation methods. It was conducted in a special laboratory which was designed and instrumented to gain access to the internal processes of competing teams of people. The principal tools were a closed-circuit television network, a voice-listening and recording network, a telephone switchboard and associated telephone and intercom network, and a variety of human observers. This study involved arranging the closed-circuit television and sound network so that a team could request picture and sound of one of the other four teams to be displayed on a monitor in its own game room and to have the associated verbal behavior of the observed team switched to earphones in its game room. This was accomplished by calling a central switchboard, at which the telephone operator switched the appropriate picture and sound as requested. A nominal charge of $1,000 was assessed against the accumulated earnings of the requesting team whenever a request was processed by the telephone operator. The charge was nominal since a team could make a profit of up to around $500,000 per session, depending upon how it fared in winning contracts through competitive bidding.

Task Environment

The Research and Development Game was played by five three-man teams, which competed to earn the most money by winning contracts. In each of twelve 3-hr. periods, the teams bid on contracts awarded for products consisting of 5-letter word-pairs. The research phase of the research and development process was introduced by converting the first word (e.g., judge) into the second word (e.g., final) according to a set of rules specifying legal and illegal intermediate five-letter words and the costs of the conversion processes. The total cost of the conversion processes thus became the manufacturing cost of the item, and the team devising the least expensive solution was in the most advantageous position for bidding on production contracts involving that item. The team would then arrive at a bid by adding to its manufacturing cost the profit per unit it hoped to make. A ceiling of twice the manufacturing cost was imposed to eliminate the fantastic profit resulting from a situation in which a team knew that it was the only bidder. Contracts were awarded to the lowest bidder.

At the beginning of each game year (a year was played in one 3-hr. laboratory period), all the teams received an identical annual contracts list of products (word-pairs) for which contracts were awarded during the next 12 mo. The list represented the possible products the team could manufacture in its hypothetical factory if it was successful in winning a bid. Each team was given an identical product specification sheet for each word-pair, listing one possible word-chain and its cost of production. A team could elect to bid on a contract for this product using the given production cost, or it could elect to try to find a less expensive solution to the word-pair. The products of the different teams were assumed to be of equal quality and different only in the cost of production.

Winning a contract scheduled a team's hypothetical factory for a 3-mo. period of compressed game time and there were limits as to how far in advance a team could schedule its factory. Since the number and size of contracts was not enough to keep all factories fully occupied during a game year, a limited number of fixed-fee contracts were available. The award of a fixed-fee contract gave the team a modest profit, but tied up the factory for a 4-mo. period. These contracts were awarded on a first-come, first-served basis to requesting teams until the supply for the year was used up.

All teams were provided with telephones, by means of which they could submit bids, request information about the bids or scheduling status of other teams, or request TV and sound access to another team. All telephone calls were charged at the rate of $1,000 per call. These charges, when added to a fixed yearly overhead cost, represented the yearly costs of operations, to be offset by profits from taking fixed-fee contracts or successful bidding on competitive contracts. Starting the game with no capital, the teams' profits and losses accumulated over each year until the game was completed. Thus, a performance ranking was available in dollars year by year, and in terms of net accumulated gain or loss.

The problem faced by the team was a rather complex one because of the many interactions. It was essentially the problem of optimal allocation of limited resources (time, skill, effort) over the full range of information processing and decision making in a rapidly changing environment. In addition, since these were self-organizing teams, they confronted the problem of devising an effective internal organization for utilizing their human resources.

RESULTS

This article concentrates on the consequences of opening up a microcosm of typical business competition by means of providing a technology for gaining mutual behavioral access. The preliminary results of this study are presented by attempting to answer a series of questions.

1. Did the teams utilize behavioral access? The answer is a resounding "yes" in the initial few sessions of the first game in which the number of requests indicated that the teams were observing between two and three other teams during the year. This usage dropped to approximately one other team during the second game and one team, the consistently poorest performer, ignored this access completely during the second game. Some teams discovered that they could establish a two-way communication network by tuning in on each other at the same time and this capability became the basis for coalition formation.

2. What were the effects of behavioral access on team behavior? The consequences of making behavioral access available to the teams were quite clear and dramatic. Although considerable time and effort was devoted to surveillance of the other teams, all the teams developed antiaccess strategies for keeping what they considered to be important information from their competitors. In some cases, rather elaborate codes were designed to exchange information within the team. A common device was to speak in a whisper on the telephone when sending in a bid or to disappear under the table with the telephone, to get away from the microphone. Noisemakers were used to mask out communications at crucial times. A second kind of antiaccess strategy consisted of holding team meetings outside the laboratory periods to permit the discussion of plans, strategies, and tactics in a place secure from the technological spy. The teams varied in their reaction to mutual access, ranging all the way from excessive concern with secrecy on the part of the teams in the middle range of performance, to less concern on the part of the winning team and the consistently losing team. The consistently losing team decided to ignore the spy capability completely during the second game and to invest all of its resources in trying to obtain better solutions to the word-chains. The consistently winning team in the first game developed a strategy of utilizing their substantially lowered cost on a word-chain once for a big killing and then broadcasting the improved solution to all listeners in the expectation of reducing profit for the other teams when that contract came up for bidding again. In summary, simultaneous mutual behavioral access appears to have been viewed as an intolerable condition under which to conduct typical business competition. The attitudes toward competition in these teams seemed to assume the necessity for covert plans, strategies, and actions. In the present treatment, the teams were uncertain about whether they were being observed (except in the two-way case described above) and his uncertainty undoubtedly contributed to the general level of paranoid behavior exhibited.

3. Under what conditions might mutual behavioral access become more tolerable as a condition for competition? It may be hypothesized that the timing of access is one important variable. Storage, for example, on TV tape of the behavioral data would permit mutual access with delay. A progressively shorter delay might be utilized in a learning design as

the participants progressively adapted to the condition. We propose to conduct investigations on the consequences of making available mutual behavioral access to the previous session during a given period of operations and to search for other variables which might lead to a progressive elimination of the strategies of secrecy.

4. What are some possible implications of experimentation on mutual behavioral access? It is possible that mutual behavioral access, as a condition for organizational, national, and international development, may become a powerful tool for finding the minimum level of cooperation required to stabilize competition during periods of explosive cultural change. The technology for obtaining, storing, and disseminating the essential behavioral data is now well developed and should be applied to a variety of competitive situations. In particular, for example, we could start to store high-level planning and decision-making meetings of our national government elite and encourage other nations to do so. With these essential behavioral documents in storage, we could engage in negotiations for exchange, with whatever delay would be mutually agreeable. Mutual behavioral access and the feedback from observing how plans and strategies are followed thus become the basis for the gradual development of a minimal level of trust and understanding between competing entities.

(Received November 5, 1969)

ournal of Applied Psychology
971, Vol. 55, No. 1, 46–49

THE SYSTEM APPROACH:

A PRELIMINARY EXPLORATORY STUDY OF THE RELATION BETWEEN TEAM COMPOSITION AND FINANCIAL PERFORMANCE IN BUSINESS GAMES[1]

JOHN L. KENNEDY[2]

Department of Management, Graduate School of Business Administration
University of Southern California

Sixteen 3–5-man self-organizing teams, composed to vary on a dimension of conceptual complexity of the individual members, competed in a laboratory-based business game. Competitive performance was measured in terms of profit or loss accumulated over a 12-wk. period of operations. Variation in team size (3–5 members) was not related to performance. The primary composition variable, conceptual complexity, as measured by a Paragraph Completion Test, was significantly related to financial performance.

A previous article (Kennedy, 1962) described several laboratory studies of miniature organizations ranging in size from 30–40 members (Air Defense crews) to three-man management teams. These interrelated systems of people may be referred to as organizations or teams as opposed to groups, because they worked together on a common task over many sessions and achieved a relatively stable structure through the processes of organizational development.

The study of development of the management teams was made possible by offering a regular undergraduate course, "Human Problems in Industry," the laboratory portion of which consisted of weekly periods of competition between teams in a simulated business environment or game during a 3-mo. organizational life. The present article reports the results of two such studies, the first conducted during the fall term, 1964, and the second during the fall term, 1965, in which the composition of the miniature organizations was varied systematically and the effects of the several types of compositions were evaluated on the basis of team profit or loss as the measure of performance.

Team Composition Variables

The teams in both years were composed of Princeton students, 35 in 1964 and 26[3] in 1965. The three composition variables were: (*a*) score on a Paragraph Completion Test (Schroder, Driver, & Streufert, 1967), a projective test of cognitive complexity; (*b*) number of participants in each team, and (*c*) college class.

The variable of cognitive complexity is one measure of the level of maturity of human information processing capability. The Paragraph Completion Test yields a score, from 0–14, which purports to measure the individual's characteristic ways of perceiving stimuli, that is, how he differentiates between stimuli to create classes of people and objects, and how he integrates these experiences into higher order concepts. At the low end of the scale will be found the people with "black-white, I like-I don't like" kinds of differentiations and simple integrative concepts. At the high end of the scale are the more sophisticated differentiators and those whose attitudes and concepts are more complex. Table 1 shows the distribution of scores of participants in the 1964 and 1965 studies.

As indicated in Table 1, the distributions were arbitrarily dichotomized (between Scores 5 and 6) into abstract individuals and concrete individuals and teams were composed by varying the percentage of abstract members. The eight 1964 teams consisted of three five-man teams and five four-man teams; the eight 1965 teams consisted of two four-man teams and six three-man teams. The variable of college class (a rough measure of chronological age and maturity) was used to attempt to put at least one senior on each team to help the team solve the leadership problem; the variable was also used to distribute team membership over the college classes as evenly as possible. Since there were 11 seniors, 13 juniors, 7 sophomores, and 4 freshmen in the 1964 class, and 6 seniors, 6 juniors, 5 sophomores, and 4 freshmen in the 1965 class, exactly even balancing could not be achieved.

---

[1] This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to the author, Department of Management, University of Southern California, University Park, Los Angeles, California 90007.

[3] Five Woodrow Wilson School mid-career fellows, ranging in age from 35–45 yr., also participated in the 1965 game. One team was composed entirely of those men and another team had a single member from this older group.

Table 2 describes the compositions achieved in 1964 and shows the financial performance of the teams in terms of profit or loss over the first five 2-hr. periods of game operations. The data on the eight teams are presented in two groups or sections of four teams (orange section and black section) since only four teams played at any one time. Table 3 presents the same information for the first five sessions of the 1965 teams. After the first five sessions of the 1965 game, some personnel were exchanged between teams so that the 11-session data of 1965 teams are not comparable to similar data of 1964 teams.

TABLE 1

DISTRIBUTIONS OF SCORES ON PARAGRAPH COMPLETION TEST

Paragraph Completion Test score	1964 participants	1965 participants
Concrete		
0		
1		
2		2
3	5	
4	10	5
5	10	6
Abstract		
6	6	9
7	3	2
8	2	2
9	1	
10		
11		
12		
13		
14		

TABLE 2

COMPOSITION AND FINANCIAL PERFORMANCE—1964 TEAMS

Team	N	College class	% abstract members	Profit or loss for first 5 sessions	Profit or loss for 11 sessions
Black section					
C	5	1 Sr. 2 Jr. 1 Soph. 1 Fresh.	60	1,945,674	6,244,525
D	4	1 Sr. 1 Jr. 1 Soph. 1 Fresh.	75	−3,244,232	2,998,329
A	4	1 Sr. 1 Jr. 2 Soph.	0	−3,939,550	−9,967,752
B	5	1 Sr. 2 Jr. 1 Soph. 1 Fresh.	0	−5,169,864	−5,140,701
Orange section					
N	4	1 Sr. 2 Jr. 1 Fresh.	75	−2,623,681	689,620
S	4	3 Sr. 1 Soph.	0	−4,523,597	−6,168,780
R	5	1 Sr. 2 Jr. 1 Soph. 1 Fresh.	0	−4,585,509	−6,270,209
M	4	1 Sr. 2 Jr. 1 Soph.	25	−10,401,654	−8,354,613

Note.—Sr. = senior, Jr. = junior, Soph. = sophomore, Fresh. = freshman.

Synthetic Environments

The present studies were carried out in the Social Psychology laboratory of the Department of Psychology at Princeton University. A section of four teams played a business game[4] for a 2-hr. period, once a week, while

---

[4] DATA was the code name for the 1965 business game and BIG was the 1964 game. The major difference between these games lay in the size of the initial resources made available to each team: $8,000,000 for DATA teams, and $4,000,000 for BIG

TABLE 3

Composition and Financial Performance—1965 Teams

Team	*N*	College class	% abstract members	Profit or loss for first 5 sessions
Black section				
C	3	1 Sr. 1 Jr. 1 Fresh.	33	22,031,000
A	4	Adults	100	− 2,004,400
D	3	1 Adult 1 Soph. 1 Fresh.	33	−10,819,000
B	3	1 Sr.	33	−23,974,260
Orange section				
M	4	2 Sr. 2 Jr.	100	57,363,000
R	3	1 Sr. 1 Jr. 1 Soph.	0	28,804,400
N	3	1 Jr. 1 Soph. 1 Fresh.	67	20,487,400
S	3	1 Sr. 2 Soph.	0	−14,782,400

Note.—Sr. = senior, Jr. = junior, Soph. = sophomore, Fresh. = freshman.

the members of the other section observed and listened to the verbal interaction through one-way mirrors. On a later afternoon in the week, the first observing section played and the initial playing section observed.

At the first meeting of the course, each participant was given a copy of the player's manual to read and assimilate. Testing with the Paragraph Completion Test was carried out at this time. The teams were designed before the second meeting of the first week and, during the second meeting, each four-team section was given a 1½-hr. dry run of experience with the game. Thereafter, each section played a 2-hr. game "year" per week.

TABLE 4

Team Size Related to Financial Performance

No. of members	No. of teams	Average section rank	Average profit or loss in first five sessions[a]	Average profit or loss for 11 sessions of BIG & 5 sessions of DATA[a]
3	6	2.83	3,624,523	3,625,500
4	7	2.14	5,520,841	4,936,485
5	3	2.67	−2,603,233	−1,722,128

[a] See Tables 2 and 3.

RESULTS

The principal results of these studies are exhibited in Tables 2 and 3. By reference to the composition and performance data contained in these tables, it is possible to determine the consequences of financial performance of (*a*) varying the size of the teams, and (*b*) varying the percentage of abstract members of the teams.

Team Size

Table 4 shows the average section rank in profit or loss and the average profit or loss for the first five sessions for teams with three, four, and five members.

TABLE 5

Cognitive Complexity Related to Team Performance

Composition	No. of teams	Average profit or loss in first 5 sessions[a]	Average profit or loss for 11 sessions of BIG and 5 sessions of DATA[a]
Above 50% abstract members	6	10,681,908	14,296,412
Below 50% abstract members	10	−2,736,043	−3,464,232

[a] See Tables 2 and 3.

---

teams. These amounts were subtracted from the net accumulated gain or loss of the teams to produce the more comparable profit and loss figures of the criterion measure of performance.

A Kruskal-Wallis one-way analysis of variance test (Siegel, 1956) applied to the data of Table 4 recommends accepting the null hypothesis of no difference in financial performance related to team size.

Percentage of Abstract Members

Table 4 summarizes the financial performance of the teams when the composition variable of percentage of abstract members is used. The Kolmogorov-Smirnov test (Siegel, 1956) yields a nonsignificant chi-square between the performance distributions of differently composed teams after the first five sessions, but a chi-square significant at the .05 level between the distributions obtained by combining the full 11-session performance of BIG teams and first five sessions of DATA teams.

At the end of the first five sessions in both BIG and DATA, a winner was declared and the teams were returned to initial conditions to play a second game. In DATA, changes in personnel of some of the teams were made before the start of the second game by exchanging members between teams. Thus, it is not possible to compare the 11-session performance in BIG with the 11-session performance in DATA. But it is possible to compare the performance of the new teams in the second half of DATA, with the results shown in Table 5. Table 6 makes the same breakdown of these data.

The distributions are not significantly different because of the small *N*, but the average difference between teams composed on the cognitive complexity variable is about the same order of magnitude as in Table 5 and in the same direction.

TABLE 6

COGNITIVE COMPLEXITY RELATED TO TEAM PERFORMANCE

Composition	No. of teams	Average profit or loss for second 6 sessions of DATA
Above 50% abstract members	3	−3,983,500
Below 50% abstract members	5	−12,799,040

DISCUSSION

The discovery of a statistically significant difference between the performance of teams varying in percentage of abstract members is rather remarkable for several reasons:

1. The arbitrary assignment of the participants to concrete and abstract categories was based on small differences. Persons scoring 5 and below were classified concrete and those scoring 6 and above were classified abstract. A replication of this study should utilize participants scoring at more extreme levels on the cognitive complexity scale.

2. The small number of teams and the high variability of team performance is also a factor. Within the limitations of a preliminary study, however, the data reported here support the conclusion that the composition variable of cognitive complexity, as measured by the Paragraph Completion Test, is related to team performance when team compositions of below 50% abstract members are compared with above 50% abstract members. In addition, there is some suggestion that team performance differences become larger as the number of sessions increases, that is, the effect of the composition variable increases over time and sessions.

REFERENCES

KENNEDY, J. L. The system approach: Organizational development. *Human Factors,* 1962, **4**, 25–52.

SCHRODER, H. M., DRIVER, J. J., & STREUFERT, S. *Human information processing.* New York: Holt, Rinehart and Winston, 1967.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

(Received November 5, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 50–56

COMPARABILITY OF OLD AND REVISED FORMS OF THE STRONG VOCATIONAL INTEREST BLANK FOR WOMEN[1]

RICHARD W. JOHNSON[2]

Counseling Center, University of Wisconsin

The old and revised forms of the Strong Vocational Interest Blank for Women (SVIB-W) were administered to a volunteer sample of 124 female *S*s. The scores on the 29 occupational scales common to both forms of the SVIB-W were highly related (*Mdn.* $r = .83$) for most of the scales. The level of scoring on some of the scales changed markedly, however, largely dependent upon the year of sampling of the original occupational group members. Because of the latter, correlations between individual profiles for the two forms were only moderately high (*Mdn.* $r = .71$). Conclusions based upon profile analyses of the old form should not be generalized to the new form. A short form of the Basic Interest Scales, scorable on the old SVIB-W, may serve as a bridge between the old and new SVIB-W for use in longitudinal research projects.

The 1969 revision (Form TW398) of the Strong Vocational Interest Blank for Women (SVIB-W) contains a number of major changes (Campbell, 1969). The number of scales has been greatly increased, many of the items have been changed, new criterion samples have been obtained for most of the occupational scales, and a new women-in-general (WIG) reference group has been established. In each instance, the changes are more extensive than those made with the 1966 revision of the men's (SVIB-M) form (Campbell, 1966).

Most of the information provided by the 1946 version (Form W) of the SVIB-W (Strong, 1946) can still be obtained with the revised form. All but 4 of the 33 occupational scales and both of the nonoccupational scales found on the old form have been restandardized. In addition, 29 new occupational scales, 2 new nonoccupational scales, 19 new basic interest scales, and 6 new administrative indexes are now listed on the profile sheet.

Because of the widespread changes, the test user will want to know to what extent scores on the new form are comparable to scores on the old form. Can research findings or clinical skills based on the old SVIB-W be generalized to the new form? The purpose of this study was to compare test results obtained in testing a common group of *S*s with the old and revised forms of the SVIB-W.

METHOD

Subjects. A total of 124 female *S*s completed both forms of the SVIB-W. The *S*s ranged in age from 15 to 66 ($\bar{X} = 25.5$, $SD = 8.0$). Of the 124 *S*s, there were 42 gainfully employed adults, 7 housewives, 16 graduate students, 58 undergraduate college students, and 1 high school student. The *S*s were drawn from a wide variety of fields. Among the employed *S*s, 14 were in the field of education, 10 in a health-related field, 12 in secretarial or office work, and 6 in other fields. The students were concentrated in pharmacy ($N = 33$) or education ($N = 19$). Most of the remaining students were enrolled in the College of Letters and Sciences. All *S*s were volunteers.

Procedures. The two forms of the SVIB-W were administered in counterbalanced order. The *S*s were instructed to take the forms in the order indicated within 24 hr. of each other. Most *S*s completed the test booklets at home. Test profiles for *S*s taking the SVIB at home are essentially the same as profiles obtained under standard conditions at test administration centers (Stephenson, Smith, & Rimland, 1965). The *S*s were mailed copies of their test profiles together with a brief interpretive pamphlet at the conclusion of the study.

Statistical analysis. Scale comparabilities for the two forms of the SVIB-W were determined by correlating all of the profile scores on the old form of

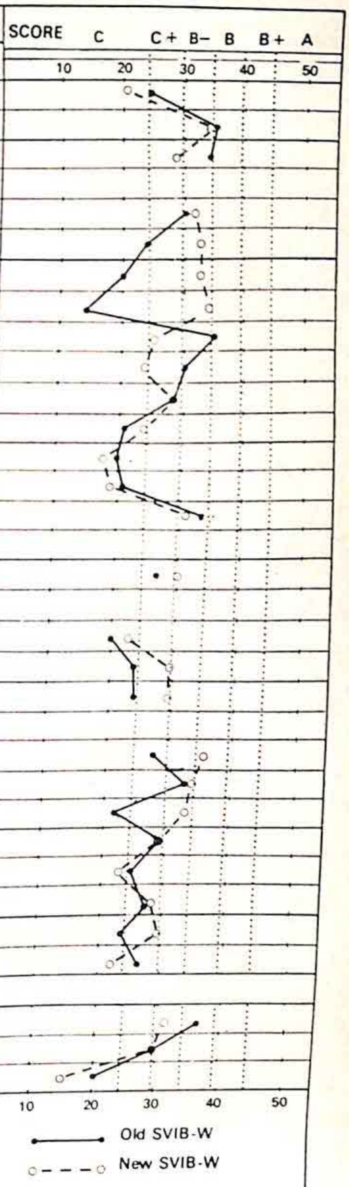
---

[1] Use of the University of Wisconsin Computing Center was made possible by support, in part, from the National Science Foundation and the Wisconsin Alumni Research Foundation through the University of Wisconsin Research Committee.

This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to the author, Counseling Center, University of Wisconsin, 415 West Gilman Street, Madison, Wisconsin 53706.

GROUP	OCCUPATIONAL SCALE	DATE OF TESTING ORIGINAL CRITERION GROUP	r	MEAN old	MEAN new	SD old	SD new
I	MUSIC TEACHER	1952*	83	25	21	14	13
	MUSIC PERFORMER	1952*	82	35	34	13	13
II	ARTIST	1934, 1941	91	34	28	10	12
	AUTHOR						
	LIBRARIAN	1934	67	30	32	8	12
	ENGLISH TEACHER	1934	83	24	33	14	15
III	SOCIAL SCIENCE TEACHER	1934, 1941	70	20	33	12	12
	Y W C A SECRETARY	1934	63	14	34	10	13
	SOCIAL WORKER	1953*	79	35	25	13	13
	PSYCHOLOGIST	1942	87	30	24	12	14
	LAWYER	1934, 1941	89	28	28	11	12
IV	LIFE INSUR SALESWOMAN	1937, 1941	86	21	24	10	11
V	BUYER	1942*	76	20	18	12	10
	BUSINESS EDUCATION TEACHER	c 1941	84	21	19	10	11
	STENOGRAPHER-SECRETARY	1934, 1947	85	33	31	10	12
	OFFICE WORKER						
VI	ELEMENTARY TEACHER	c 1939	83	27	30	10	11
	HOUSEWIFE						
	HOME ECONOMICS TEACHER	1941	74	20	23	12	12
	DIETITIAN	1941	73	23	29	12	11
VII	PHYS ED TEACHER (H S)	1939	66	24	29	9	13
	PHYS ED TEACHER (COLL)						
	OCCUPATIONAL THERAPIST	1942	69	28	36	12	12
	PHYSICAL THERAPIST	1957	86	33	34	13	14
	NURSE	1934	70	22	33	10	11
VIII	PHYSICIAN	1934, 1941	88	29	28	12	15
	DENTIST	1934, 1942*	88	25	23	13	12
	LABORATORY TECHNICIAN	1941	90	27	28	14	17
IX	MATH-SCIENCE TEACHER	1934, 1941	90	24	29	13	13
	ENGINEER	1959*	87	26	22	13	15
SUPPLEMENTAL SCALES							
	SPEECH PATHOLOGIST	1965*	87	37	32	14	13
	COMPUTER PROGRAMMER	1964, 1966*	92	29	29	13	14
	CATHOLIC SISTER TEACHER	1961*	70	20	15	11	11
NON-OCCUPATIONAL SCALES							
	MASCULINITY - FEMININITY	1934	70	48	46	11	11
	ACADEMIC ACHIEVEMENT	1961*	95	48	48	11	11



* Original criterion group used in 1969 revision. All other criterion groups used in the revision were tested in 1965, 1966 or 1967 (Campbell, in press).

FIG. 1. Comparison of scores on old and new SVIB-W scales (N = 124). (The scales are listed as they appeared on the old version of the inventory. The names of some of the scales have been changed slightly in the revision, e.g., Nurse has been changed to Registered Nurse and Laboratory Technician has been changed to Medical Technologist. The four scales with no data were dropped from the revised form.)

the SVIB-W with all of the profile scores on the new form. Of particular interest were the correlations between the 29 occupational and 2 nonoccupational scales with identical or similar names on the old and new forms.

Profile comparability was determined separately for each *S*. The scores obtained by each *S* on the 29 similar scales on the old form were correlated with the scores obtained on these same 29 scales on the new form.

Scores on the new Basic Interest Scales were correlated with scores on a short form of the Basic Interest Scales available for the 1946 SVIB-W.[3] The

---

[3] Information on procedures for scoring old SVIB-W (Form W) answer sheets (pre-1969) on

TABLE 1

HIGHEST POSITIVE CORRELATION COEFFICIENTS BETWEEN OLD SCALES DELETED FROM SVIB-W AND NEW SCALES

Old scale	$\bar{X}$	SD	New scale	$\bar{X}$	SD	r
Author	33	11	Artist	28	12	.84
			Newswoman	33	15	.82
Office Worker	29	11	Instrument Assembler	23	13	.86
			Bankwoman	25	11	.85
			Telephone Operator	20	14	.85
Housewife	31	9	Saleswoman	19	12	.82
			Telephone Operator	20	14	.79
Physical Education Teacher (College)	24	9	Occupational Therapist	36	12	.71
			Physical Therapist	34	14	.64

Note.—$N = 124$.

Basic Interest Scales on the revised form include all of the items found on the old form plus new items which are included only in the revised form of the SVIB-W booklet.

All correlation coefficients were Pearson product-moment rs. All statistical analyses were performed on the University of Wisconsin, Control Data Corporation 3600 computer.

RESULTS

The median correlation coefficient between the old and revised occupational scales was .83 (see Figure 1). The intercorrelations ranged from a low of .63 (YWCA Secretary) to a high of .92 (Computer Programmer).

Scores for 20 of the 29 new scales were based on new occupational criterion groups. The median old-new intercorrelation for the 20 scales based on new criterion groups was identical (.83 in both cases) to the median r for the 9 occupational scales based on old criterion groups.

The academic achievement (AACH) scale was only slightly modified by dropping 8 of the 55 items which were scored on this scale. The high intercorrelation ($r = .95$) between the old and revised AACH scale suggests that effect of this change was negligible. The new femininity-masculinity scale (FMII), however, has been substantially changed. This scale now contrasts the interests of equal samples of men and women employed in 18 different occupations. Unemployed women are no longer included in the sample. Only items common to the men's and women's forms of the SVIB are scored. As a result of these changes, the correlation between the old and new forms was only moderately high ($r = .70$).

The intercorrelations were almost always higher with a similar scale than with any other scale. There were only a few cases in which a key was more highly correlated ($p < .05$) with a nonsimilar key than a similar key. The old YWCA Secretary key was more highly correlated with the new Guidance Counselor key ($r = .76$) than with the new YWCA Staff member key ($r = .63$). The new YWCA Staff member key, on the other hand, was more closely related to the old Social Worker key ($r = .77$). The new Librarian key was more highly correlated with the old English Teacher scale ($r = .80$) than with the old Librarian scale ($r = .67$).

Despite the relatively high intercorrelation of the old and new occupational scales, the level of scoring on some of the scales varied considerably. Of the 20 similar scales based upon new occupational groups, 10 differed by at least 5 standard scores or more. In all but 2 of these 10 cases (Artist, Psychologist) the scores were higher on the new form. The differences were most notable on the very old scales. Of the four scales standardized in 1934, three (YWCA Secretary, Nurse, English Teacher) showed increases in scores of nearly one standard deviation or greater. Scores on the YWCA key moved from lowest (standard score = 14) for the group on

---

the short form of the Basic Interest Scales may be obtained from David P. Campbell, Director, Center for Interest Measurement Research, University of Minnesota, Minneapolis, Minnesota 55455.

the old form to tied for second highest (standard score = 34) on the new form. In contrast with these results, *S*s scored lower (or equal) on the new form on all nine occupational scales based upon old criterion groups. In three cases (Social Worker, Catholic Sister Teacher, and Speech Pathologist), the difference was equal to or greater than 5 standard score points.

Of the four old SVIB-W scales which have not been replaced on the new form, all are highly correlated with a new scale except Physical Education Teacher (College) (see Table 1). Inspection of both correlation coefficients and mean scores indicates that the Newswoman scale is a reasonable substitute for the old Author scale and that the Bankwoman scale closely resembles the old Office Worker scale. Relative scores on the Housewife key may be estimated with a fair degree of accuracy by scores on some of the nonprofessional keys, particularly Saleswoman and Telephone Operator. However, *S*s in this study scored quite a bit lower on these scales than on the old Housewife scale. No new scale corresponds closely with the old Physical Education Teacher (College) scale. The new Physical Education Teacher scale, based upon the responses of high school physical education teachers, does not replace the old scale for college teachers ($r = .47$).

Correlations between the new nonsimilar occupational keys and old SVIB scales are shown in Table 2. Only the highest positive correlation coefficients are reported. The following scales appear to be most successful in

TABLE 2

HIGHEST POSITIVE CORRELATION COEFFICIENTS BETWEEN NEW SCALES ADDED TO SVIB-W AND OLD SCALES

New scale	$\bar{X}$	*SD*	Old scale	$\bar{X}$	*SD*	r
Entertainer	30	14	Music Performer	35	13	.81
Model	31	12	Life Insurance Saleswoman	21	10	.63
Art Teacher	24	16	English Teacher	24	14	.62
Interior Decorator	20	14	Author	33	11	.66
Newswoman	33	15	Author	33	11	.82
Language Teacher	35	15	English Teacher	24	14	.79
Recreation Leader	34	11	Music Teacher	25	14	.59
Director, Christian Education	21	13	Music Teacher	25	14	.77
Guidance Counselor	29	12	Social Worker	35	13	.79
Translator	33	13	Artist	34	10	.71
Chemist	14	15	Physician	29	12	.78
Mathematician	18	14	Physician	29	12	.80
Army-Enlisted	27	9	Math-Science Teacher	24	13	.53
Navy-Enlisted	31	9	Laboratory Technician	27	14	.68
Army-Officer	32	10	Lawyer	28	11	.80
Navy-Officer	38	8	Lawyer	28	11	.59
Accountant	21	11	Computer Programmer	29	13	.57
Bankwoman	25	11	Office Worker	29	11	.85
			Business Education Teacher	21	10	.85
Public Health Nurse	32	12	Nurse	22	10	.51
Licensed Practical Nurse	20	12	Dietitian	23	12	.73
			Nurse	22	10	.73
Radiologic Technologist	31	15	Laboratory Technician	27	14	.86
Dental Assistant	25	14	Dietitian	23	12	.78
Executive Housekeeper	22	11	Dietitian	23	12	.78
Saleswoman	19	12	Office Worker	29	11	.83
Telephone Operator	20	14	Office Worker	29	11	.85
Instrument Assembler	23	13	Office Worker	29	11	.86
Sewing Machine Operator	13	13	Office Worker	29	11	.83
Beautician	30	10	Buyer	20	12	.76
Airline Stewardess	31	13	Stenographer-Secretary	33	10	.52

Note.—N = 124.

TABLE 3

FREQUENCY DISTRIBUTION OF PRODUCT-MOMENT CORRELATION COEFFICIENTS BETWEEN OLD AND NEW SVIB-W PROFILES

r	f	Σf	Cumulative %
.90–.94	3	124	100
.85–.89	14	121	98
.80–.84	21	107	86
.75–.79	14	86	69
.70–.74	15	72	58
.65–.69	9	57	46
.60–.64	9	48	39
.55–.59	12	39	31
.50–.54	7	27	22
.45–.49	7	20	16
.40–.44	7	13	10
.35–.39	3	6	5
.30–.34	2	3	2
.25–.29	1	1	1

tapping interests not adequately measured by the old SVIB ($r < .80$): Public Health Nurse, Airline Stewardess, Army-Enlisted, Accountant, Navy Officer, Recreation Leader, Art Teacher, Model, Interior Decorator, Navy-Enlisted, Translator, Licensed Practical Nurse, Beautician, Director of Christian Education, Chemist, Dental Assistant, Executive Housekeeper, Guidance Counselor, and Language Teacher. All of the other new scales correlated highly with at least one old SVIB-W scale. The new nonprofessional keys appear to be most closely related to the old business keys.

The median intercorrelation between the old and new individual profiles for the 29 similar occupational scales was .71. Although some of the interprofile correlations were relatively low (see Table 3), all but five of the rs were significantly different from chance ($p < .05$).

The relationship between the old and new group profiles (as shown in Figure 1) was statistically insignificant ($r = .31$, $p > .05$). The group profiles were less highly intercorrelated than most of the individual profiles presumably due to the restricted range of scores on the group profiles. The standard deviation of scores for most individual profiles was 10–15 points; the standard deviation of scores for the group profiles was less than 6 points in both cases.

The Basic Interest Scales for the two forms were more closely related than the occupational scales. As shown in Table 4, the median intercorrelation was .90 with a range from .86 to .95.

DISCUSSION

The median r of .83 between the old and new similar occupational scales was surprisingly high. This r is virtually identical with the r of .84 obtained between the old and new occupational scales on the SVIB-M (Campbell, 1966, p. 55; Williams, Kirk, & Frank, 1968). Because of the use of a completely new women-in-general group, new occupational groups, and new items, the relationship between the scales was expected to be lower. Furthermore, because new items were used, the test scores were based upon two separate administrations of the interest inventories. In both of the SVIB-M comparability studies, the test scores for the old and new scales were based upon a single administration of the old form of the SVIB (which included all of the items scored on the new form).

For the majority of the interest scales, S's relative performance on the new scales may

TABLE 4

CORRELATIONS BETWEEN OLD (SHORT) AND NEW (LONG) VERSIONS OF BASIC INTEREST SCALES FOR SVIB-W

Scale	Old form			New form			
	No. of items	$\bar{X}$	SD	No. of items	$\bar{X}$	SD	r
Public speaking	12	52	9	13	52	10	.90
Law/Politics	9	53	11	10	53	11	.87
Merchandising	12	49	9	13	48	9	.89
Office Practices	11	47	8	18	47	8	.92
Numbers	7	49	9	8	48	10	.93
Physical Science	10	49	11	15	49	11	.93
Mechanical	10	48	10	16	47	10	.93
Outdoors	6	49	10	9	49	10	.90
Biological Sciences	6	51	10	7	52	11	.94
Medical Service	10	51	10	17	50	10	.92
Teaching	11	50	10	12	50	10	.90
Social Service	6	51	8	8	50	9	.86
Sports	9	50	10	12	50	10	.93
Homemaking	10	49	8	11	50	9	.89
Religious Activities	5	44	10	9	44	9	.90
Music	5	51	10	6	51	10	.90
Art	7	53	9	9	52	9	.86
Performing Arts	11	53	10	13	52	10	.93
Writing	12	52	10	14	52	10	.95

Note.—$N = 124$.

be fairly accurately estimated by her performance on the old scales. Relative scores on the new YWCA Staff Member, Physical Education Teacher, Librarian, and Occupational Therapist scales are most divergent from relative scores on the old scales.

Test users will want to consider absolute level of scoring on the new scales as well as relative level of scoring. In general, there was a tendency for *S*s to score higher on new scales which formerly had been based upon old criterion groups (1934 vintage) and lower on new scales which were based on the same criterion groups as before. The levels of scoring on the YWCA Staff Member, Social Science Teacher, Nurse, Social Worker, English Teacher, and Occupational Therapist keys were most affected.

A good part of the difference in level of scoring may be attributed to the changing popularity of interest items (Campbell, 1968). As items change in popularity, scores on interest scales will go up or down depending upon year of testing of the criterion group and the WIG group. Ideally, the occupational groups and the WIG group should both represent the current interests of females. If either one is out of step with the times, the scores will be somewhat distorted.

The relatively low intercorrelations between the old and new individual profiles were somewhat surprising in view of the high correlations for the individual scales. The correlation coefficients between the individual scales, however, are not sensitive to differences in level of scoring on the two forms. If changes in level of scoring had been uniform from scale to scale, the interprofile correlations would also have run high. Because some of the scales showed a much greater variation in level of scoring than other scales, the relative positions of the occupational scores on the two forms were significantly altered in many cases. The greater the change in profile shape, the lower will be the interprofile correlation coefficient.

How do these results affect the use of the test in counseling and research? Based upon Hoyt's (1960) study, clinical judgments of profile comparability are closely related to statistical judgments of profile comparability. He found that a Spearman rank-order correlation coefficient (a close approximation of Pearson's *r*) of .65 or less between SVIB-M profiles indicated "at least one important difference" in interpretation. Using this criterion, nearly two-fifths (39%) of the *S*s tested in this study produced profiles with important differences between them. Presumably, counseling based upon the new profile would significantly vary from counseling based upon the old profile for a substantial number of *S*s.

Research studies based on profile analysis of the old SVIB-W may be particularly misleading. The lack of a significant correlation between the old and new group profiles suggests that conclusions based on profile analyses for groups tested in recent years with the old form of the SVIB-W should not be generalized to the new form.

The correlations between the abbreviated Basic Interest Scales found on the old form and the full-length Basic Interest Scales found on the new form are high despite the fact that relatively few items (5–12) are scored on the scales in the short version. The magnitude of the relationship suggests that the Basic Interest Scales on the old form of the SVIB-W may possibly serve as an adequate measure of interests for investigators conducting longitudinal studies with the old form of the SVIB-W.

CONCLUSION

Findings based upon the relative order of scores on individual scales on the old SVIB-W may be generalized to the new SVIB-W with some degree of accuracy for most scales. It is important to note, however, that the level of scoring on some of the scales changed markedly. The scores on some scales, especially those scales originally based upon very old criterion groups, ran considerably higher on the new form. Scores on other scales, particularly scales based upon the same criterion groups used in the 1946 form, were lower on the new form.

Because of the varying change in level of performance on the two forms, findings based upon profile analysis will be significantly altered. Clinical insights or research results based upon intraprofile comparisons for the old form should not be generalized to the new form of the SVIB-W. Counseling experience with the new form will have to be acquired and new research data will have to be collected.

For researchers midway in a longitudinal research project, the short form of the Basic Interest Scales will provide a fairly complete map of an *S*'s interests which should conform closely with the results obtained with the Basic Interest Scales on the new form.

In conclusion, the new form of the SVIB-W yields information sufficiently disparate from the old form to suggest that the two forms are not equivalent for most purposes. While the validity of the new form is still relatively unknown, most users will prefer the new form because of the use of current occupational criterion groups in most cases and the addition of new scales tapping new fields of interest. The recency of the development of the scales appears to be critical. It was the scales originally based on the oldest criterion groups which produced the most discrepant results when revised.

REFERENCES

CAMPBELL, D. P. *Manual for Strong Vocational Interest Blanks for Men and Women.* Stanford: Stanford University Press, 1966.

CAMPBELL, D. P. Changing patterns of interest within the American society. *Measurement and Evaluation in Guidance,* 1968, **1**, 36–49.

CAMPBELL, D. P. *1969 supplement to the manual for Strong Vocational Interest Blanks.* Stanford: Stanford University Press, 1969.

CAMPBELL, D. P. *The handbook for the Strong Vocational Interest Blank.* Stanford: Stanford University Press, in press.

HOYT, D. P. Measurement and prediction of the permanence of interests. In W. L. Layton (Ed.), *The Strong Vocational Interest Blank research and uses.* Minneapolis: University of Minnesota Press, 1960.

STRONG, E. K., JR. *Strong Vocational Interest Blank for Women (Form W).* Stanford: Stanford University Press, 1946.

STEPHENSON, R. R., SMITH, L. H., & RIMLAND, B. A comparison of responses to a vocational test taken under standard conditions at recruiting stations and responses to the same test taken as a self-administered test at home. *Research Memorandum SRM 65-3.* San Diego, California: U. S. Naval Personnel Research Activity, 1965.

WILLIAMS, P. A., KIRK, B. A., & FRANK, A. C. New men's SVIB: Comparison with the old. *Journal of Counseling Psychology,* 1968, **15**, 287–294.

(Received November 5, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 57–65

JOB CONTENT AND CONTEXT FACTORS RELATED TO SATISFACTION FOR DIFFERENT OCCUPATIONAL LEVELS [1]

THOMAS B. ARMSTRONG [2]

Teachers College, Columbia University

Herzberg's theory was tested using Darley and Hagenah's rationale relative to occupational level. Ratings of satisfaction and importance for the job content and context factors and overall job satisfaction were obtained from 200 engineers and 153 assemblers. The proposed job factor dichotomy was not supported. However, satisfaction with the content factors made the greatest contribution to overall job satisfaction, regardless of occupational level. Conversely, ratings of job factor importance were a function of occupational level; content aspects were most important for engineers, and context for assemblers. Several demographic variables failed to influence the findings. The conclusion is that both theories represent oversimplifications.

Recent job satisfaction research has been strongly influenced by the theory proposed by Herzberg, Mausner, and Snyderman (1959), which is based on a content-context job factor dichotomy. The main hypothesis states that favorable feelings toward content factors, like achievement, contribute primarily to overall job satisfaction, but do not generally contribute to job dissatisfaction per se. Hence, content factors act mainly as satisfiers. Conversely, positive feelings toward context aspects, like job security, contribute to neutralizing dissatisfaction, but do not generally contribute to job satisfaction per se. Thus, context factors act mainly as dissatisfiers. Moreover, these diverse and curvilinear relationships are asserted to hold regardless of occupational level.

The results of research directed at the two-factor theory have been mixed. In essence, when Herzberg's method involving recall, or the critical incident technique, has been employed, the theory has generally been supported (Myers, 1964; Schwartz, Jenusaitis, & Stark, 1963). Conversely, when other methods have been used, the theory has not typically been upheld (Ewen, 1964; Hinrichs & Mischkind, 1967; Lindsay, Marks, & Gorlow, 1967).

Regardless of method, the effect of occupational level is evident in the research and has been documented by Burke (1966) and House & Wigdor (1967). Specifically, Myers (1964) followed Herzberg's method and the results showed that the frequency of appearance for the two types of job factors differed for his professionals (engineers) and semiskilled workers (assemblers). Malinovsky and Barry (1965), using a correlation approach with a satisfaction questionnaire for a low-level occupation, maintenance and watchmen, reported that the two types of factors exhibited equivalent relationships with general job satisfaction. Friedlander (1965a) factor analyzed an importance questionnaire and found that the value of the two kinds of factors differed as a function of occupational level and subgroup.

On the other hand, it is surprising to find that a related, but complex, question has been relatively neglected in surveying this literature. To what extent, if any, were the results of studies using high- and low-level occupations affected by the dimension used to assess the job factors, satisfaction versus importance? Furthermore, if there was an effect, what are the theoretical and practical implications?

---

[1] This study is based on a doctoral dissertation submitted to Teachers College, Columbia University in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Armstrong, 1969). The author wishes to acknowledge the guidance of his committee, Albert S. Thompson, Chairman; Donald E. Super; and Robert L. Thorndike.

Appreciation is expressed to Walter J. Dick, James H. McMahon, and M. Claire Ferrono who provided, respectively, statistical consultation and assistance in data processing.

This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to the author who is now at Systems Manufacturing Division, IBM Corporation, Kingston, New York 12401.

The relationship between ratings of satisfaction and of importance for various job factors has been studied only infrequently in job satisfaction research. However, available data indicate that the two sets of ratings represent two different attitudinal dimensions (Friedlander, 1965b; Froehlich, 1958). Relative to Herzberg's theory, the bulk of studies that used satisfaction ratings for the factors (Halpern, 1966; Hulin & Smith, 1967; Wernimont, 1966) revealed that the job content type generally exhibited higher relationships with overall job satisfaction than did the job context kind. However, as noted earlier (Malinovsky & Barry, 1965), these results may have been a function of the high-level occupations examined. Conversely, no such consistent findings stemmed from studies which used importance ratings for the job factors (Centers & Bugental, 1966; Fine & Dickman, 1962; Friedlander, 1965a). Moreover, the results of two of these studies were contradictory, and both used high- and low-level occupations. Specifically, Fine and Dickman found that, regardless of occupational level, the job content factors were judged most important in terms of both job satisfaction and performance. In contrast, Centers and Bugental reported that the value of the two types of factors was a function of occupational level in terms of keeping one's job. That is, content factors were most valued at higher occupational levels, and context kinds at lower levels.

In line with the major question raised earlier, a review of the research on Herzberg's theory revealed that no job satisfaction study used both a high- and low-level occupation and examined the two kinds of job factors on both the satisfaction and importance dimensions.

The goal of the present study was to help fill this research gap. In so doing, the rationale of Darley and Hagenah (1955) relative to occupational level was used to test the main hypothesis of Herzberg's theory. Darley's more traditional two-factor theory was drawn upon because it stipulates that occupational level indicates the type of job factor involved in job satisfaction. Simply stated, feelings about the job are derived from content factors for high-level occupations, and from context aspects for low-level groups. However, it should be noted that the number and, in some cases, the meanings of the specific job factors differ in the two theories. Also, contrary to Herzberg, advancement is classed as a context factor by Darley.

In the present study, the hypotheses concerned the relationships between satisfaction for the two kinds of job factors and overall job satisfaction. In addition, importance ratings for these same job factors were examined to shed added light on the problem.

Hypotheses

Two specific hypotheses were formulated: (*a*) For the high-level occupational group, the job content factors make a relatively greater contribution to overall job satisfaction than do the job context factors. (*b*) For the low-level occupational group, the job context factors make a relatively greater contribution to overall job satisfaction than do the job content factors.

METHOD

Procedure

Questionnaire. A pilot study was conducted to develop a questionnaire containing four sections.

1. A slightly modified version of Hoppock's (1935) Job Satisfaction Blank, named the Overall Job Satisfaction Index (OJSI), measured global job satisfaction. The initial edition of the OJSI was administered to 101 personnel *Ss*, scored, and a correlation analysis conducted. Of the four sets of items, the three exhibiting the highest relationship were retained. A test-retest reliability coefficient of .83 was obtained for a 2-wk. interval for the final edition of the OJSI.

2. A 104-item questionnaire called the Job Factor Satisfaction Scales (JFSS) was constructed to measure satisfaction with each of the 13 specific job factors examined, namely, five content factors (recognition, responsibility, achievement, advancement, and the work itself) and 8 context aspects (salary, job security, status, supervision-technical and interpersonal relations, interpersonal relations-peers, company policy and administration, and working conditions).

Wherever possible, items were written to reflect Herzberg's (Herzberg, Mausner, & Snyderman, 1959, pp. 143–146) definitions. One major exception was status, for which Super's (1957, pp. 6–8) definition relative to prestige and importance was used.

Selection of the final 52 pairs of positively and negatively phrased items was based primarily on the judgmental agreement of eight social scientists. An administration of the JFSS to the personnel sample, followed by an item intercorrelation analysis, was also used as an aid in item selection. The final edition of the JFSS contained 13 subscales, each of which was composed of four pairs of items expressed in a five-step, Likert-type format; and two major scales (Job Content, Job Context) which yielded a summed score for each set of specific factors. Two types of reliability estimates were high for both kinds of scales. For the subscales used to test the hypotheses, a mean test-retest coefficient of .84 was obtained for a 2-wk. interval, and a mean of $r = .85$ using the split-half method.

3. A 52-item questionnaire called the Job Factor Importance Scales (JFIS), consisting of the positively phrased items from the JFSS, assessed the relative importance of the same 13 specific job factors in terms of job satisfaction. The JFIS contained 13 subscales, each of which was composed of four items expressed in a five-step importance format; and two major scales, (Job Content and Job Context) which yielded summed scores for each set of specific factors. Internal consistency estimates were high for both kinds of scales. For the subscales, the mean correlation was .85. However, test-retest reliability for a 2-wk. period was, in general, only moderate. The mean correlation for the subscales was .66.

4. A group of items was included that concerned the demographic characteristics of age, time with company, and for the high-level group (professional engineers), job level and time on present job.

The total questionnaire, along with a cover letter explaining its purpose, was mailed to the professionals in June 1966, and administered on company time to the low-level group (semiskilled assemblers) in December 1966. Anonymity was guaranteed and participation voluntary.

Sample

The sample consisted of engineers and assemblers because both occupations are classified in the technology field (Moser, Dubin, & Shelsky, 1956), with engineers as professionals and assemblers at the semiskilled level. Hence, occupationally, the field was controlled while the influence of level was examined. Also, engineers were chosen because they represent one profession which exhibits a relatively high degree of job dissatisfaction (Deutsch & Shea, Inc., 1961). Both occupational groups were employed at a large electronic company's engineering and manufacturing facility in New York state.

A group of 285 employees was obtained from the laboratory population of 309 male, nonsupervisory engineers. Of this population, 79% were college graduates. Of the 285 engineers sampled, 200 (70%) returned sufficiently complete questionnaires to serve as *S*s. Of these 200 engineering *S*s, 86% were between 20 and 39 yr. old; 50% had 1–7 yr. company service. Also, by job level, they ranged from junior to staff engineer.

TABLE 1

OCCUPATIONAL LEVEL COMPARISON OF JOB FACTOR SATISFACTION SCALES SCORES FOR THE ENGINEERS AND ASSEMBLERS

Factor	Engineers		Assemblers		t
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	
Recognition	28.0	5.62	23.6	5.72	7.19*
Responsibility	28.7	5.00	25.3	3.36	5.37*
Achievement	29.3	6.66	26.5	6.32	4.02*
Advancement	26.8	6.91	19.4	6.15	10.01*
Work itself	29.1	6.99	24.3	6.36	6.71*
Job content score	141.9	24.97	119.0	22.06	9.11*
Salary	26.3	6.72	20.3	5.58	9.13*
Job security	33.5	4.34	31.3	3.57	5.21*
Status	24.9	6.06	21.1	5.60	6.08*
Supervision—Technical	26.6	6.80	27.0	4.20	.68
Supervision—Interpersonal relations	31.2	5.32	29.2	5.41	3.45*
Interpersonal relations—Peers	30.3	5.24	28.4	5.30	3.35*
Company policy and administration	27.6	5.10	28.4	4.85	1.50
Working conditions	28.7	5.09	23.6	4.24	10.24*
Job context score	229.3	27.65	209.4	24.72	7.11*

Note.—$N = 200$ for engineers, $N = 153$ for assemblers.
* $p < .01$ (two-tailed test).

A group of 185 assemblers was obtained from a population of 251 males assigned to the lowest level assembly jobs at the plant. Of this population, only 42% were high school graduates. Of the 185 assemblers sampled, 153 (76%) handed in sufficiently complete questionnaires to serve as *S*s. Of these 153 assembly *S*s, 73% were between 30 and 49 yr. of age, and 84% had between 10 and 15 yr. company service.

RESULTS

An initial analysis of the data was concerned with two points; the degree and variability of satisfaction, and the independence of the satisfaction ratings on the two types of factors. In regard to the first point, the OJSI data indicated that the engineers were generally only moderately satisfied with the job, but significantly ($p < .01$) more satisfied than the assemblers as shown by a t test ($t =$

TABLE 2

INTERCORRELATION MATRIX FOR FIFTEEN SCALES OF THE JOB FACTOR SATISFACTION SCALES WITH ENGINEERS

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Recognition															
2. Responsibility	44														
3. Achievement	46	63					Sector 1								
4. Advancement	56	48	59												
5. Work itself	39	66	78	45											
6. Job content score	70	79	88	78	83			Sector 3							
7. Salary	41	26	25	55	22	42									
8. Job security	28	12	25	34	11	28	25								
9. Status	54	63	71	63	63	79	31	23							
10. Supervision-Technical	55	42	45	44	46	58	36	13	44						
11. Supervision—Interpersonal relations	54	34	26	41	36	47	33	12	33	64					
12. Interpersonal relations—Peers	43	30	36	30	32	42	14	28	39	42	43			Sector 2	
13. Company policy and administration	23	08	28	37	07	27	27	40	31	18	04	34			
14. Working conditions	25	07	22	24	08	22	17	23	22	19	20	38	31		
15. Job context score	67	47	58	68	48	72	61	49	67	72	64	67	55	52	

Note.—Decimal points are omitted; $r = .14$, $p = .05$; $r = .18$, $p = .01$; $N = 200$.

3.47). The OJSI mean score for the engineers was 16.4 and for the assemblers 15.1. In both cases, the standard deviation was 3.5. In reference to the specific job factors, the data are presented in Table 1. From this table it will be noted that a similar satisfaction pattern was shown for the job factors. (Exceptions were supervision-technical and company policy and administration on which the two occupations were equivalently satisfied.)

In regard to the independence of the satisfaction ratings, the intercorrelations of the

TABLE 3

INTERCORRELATION MATRIX FOR FIFTEEN SCALES OF THE JOB FACTOR SATISFACTION SCALES WITH ASSEMBLERS

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Recognition															
2. Responsibility	29														
3. Achievement	53	46					Sector 1								
4. Advancement	58	29	57												
5. Work itself	42	42	80	59											
6. Job content score	74	56	88	81	85			Sector 3							
7. Salary	61	20	40	52	32	54									
8. Job security	11	−06	07	−08	01	02	−01								
9. Status	55	39	66	72	67	79	45	08							
10. Supervision—Technical	61	17	44	43	33	53	35	20	43						
11. Supervision—Interpersonal relations	52	27	41	41	27	49	36	33	38	76					
12. Interpersonal relations—Peers	35	10	34	38	31	40	27	06	42	31	30			Sector 2	
13. Company policy and administration	39	00	27	37	28	36	37	31	35	44	43	25			
14. Working conditions	31	15	23	27	24	31	23	14	30	34	31	18	32		
15. Job context score	70	26	58	63	50	71	63	36	71	75	76	58	68	54	

Note.—Decimal points are omitted; $r = .16$, $p = .05$; $r = .21$, $p = .01$; $N = 153$.

job factors of the JFSS for the engineers are shown in Table 2, and for the assemblers in Table 3.

Examination of Tables 2 and 3 indicates that the difference in the proportions of positively significant correlations ($p \leq .05$) within the two types of specific factors (Sectors 1 and 2 of the matrices) and between the two types (Sector 3) was not statistically significant for either occupation; respectively, 92% and 85% for the engineers, and 89% and 80% for assemblers. Hence, the proposed content-context dichotomy for the job factors was not empirically supported at either occupational level. On the other hand, the specific job content factors clustered together to a higher degree than did the job context type within both occupations as measured by a median test: for the engineers, $\chi^2 = 10.72$ ($p < .01$), and for the assemblers $\chi^2 = 4.29$ ($p < .05$).

In regard to the study's two predictions, the first hypothesis was accepted and the second rejected since the job content factors made a relatively greater contribution to overall job satisfaction than did the job context type, regardless of occupational level. Intraoccupationally, the differences in the mean correlations for the two types of specific job factors of the JFSS with the OJSI were as follows: in the predicted direction and statistically significant ($p < .01$) for the engineers (a mean r of .63 for content versus an r of .32 for context), and in the opposite direction than predicted for the assemblers (a mean r of .53 for content versus an r of .38 for context).

Since it was evident that the two occupations also differed on age and time with company, the possible effects of the demographic variables on the results were examined. Correlation and one-way analysis of variance showed that age, time with the company and, in addition for engineers, job level and time on present job, failed to individually influence the findings for either hypothesis.

For both occupations, a multiple-regression analysis, based upon a backward elimination procedure (Draper & Smith, 1967), revealed that composites of 3 job factors yielded about as good an estimate of overall job satisfaction as did using all 13 specific factors. For the initial and reduced regression equations, the multiple correlations were, respectively, .81 and .80 for the engineers, and .76 and .74 for assemblers. Furthermore, in both reduced composites, the two specific factors (designated by Herzberg as content) made a relatively greater contribution to the prediction of job satisfaction than did the one context factor. This is shown by the second-order correlation for each specific factor with the OJSI: work itself-.45, achievement-.38, and supervision-interpersonal relations-.16 for the engineers; and advancement-.40, achievement-.39, and supervision-interpersonal relations-.22 for the assemblers.

TABLE 4

OCCUPATIONAL LEVEL COMPARISON OF THE CORRELATIONS OF THE SPECIFIC JOB FACTORS OF THE JOB FACTOR SATISFACTION SCALES WITH THE OVERALL JOB SATISFACTION INDEX FOR THE ENGINEERS AND ASSEMBLERS

Factor	Correlations with OJSI		z value
	Engineers	Assemblers	
Recognition	.40	.46	− .68
Responsibility	.59	.24	3.99**
Achievement	.76	.64	2.20*
Advancement	.46	.64	−2.41*
Work itself	.79	.61	3.34**
Job content score	.76	.69	1.37
Salary	.18	.37	−1.90
Job security	.18	.08	.94
Status	.63	.59	.59
Supervision—Technical	.46	.47	− .12
Supervision—Interpersonal relations	.36	.47	−1.23
Interpersonal relations-Peers	.37	.31	.63
Company policy and administration	.16	.38	−2.20*
Working conditions	.13	.29	−1.55
Job context score	.52	.60	−1.08

Note.—$N = 200$ for engineers; $N = 153$ for assemblers.
* $p < .05$ (two-tailed test).
** $p < .01$ (two-tailed test).

Table 4 reveals that the individual relationships for 5 of the 13 specific job factors with the OJSI differed significantly as a function of occupational level.

Generally, the content factors were more potent in overall job satisfaction for the engineers than for the assemblers. However, advancement followed an opposite pattern since it was more salient for the assemblers. For the context factors, with the exception of company policy and administration, the relationships were basically similar across the two occupational levels.

For the satisfied and dissatisfied subsamples, the correlations between the job factors of the JFSS and the OJSI are reported in Table 5 along with those previously found for the two total samples. The *S*s who scored OJSI 14 or above composed the satisfied subsamples; those who scored OJSI 12 or below comprised the dissatisfied. The *S*s who scored exactly 13 (the neutral point) on the OJSI were excluded from analysis. Within an occupation, the correlations obtained for the subsamples were, except for their reduced magnitude, generally similar to those previously found for the total sample. In job satisfaction, both types of factors made significant contributions, but generally only those for content were statistically significant in job dissatisfaction. Moreover, in both job satisfaction and dissatisfaction, the typical contribution for the content type was relatively greater than context, regardless of occupational level.

Table 6 presents data on the judged importance of the job factors for the two occupational levels. It was evident that the ratings were generally restricted to the high importance end of the JFIS scale. However, the difference in the mean scores of the two major scales (Job Content and Job Context) was statistically significant ($p < .01$) within each occupation as examined by paired t tests ($t = 12.17$ for engineers and $t = 9.25$ for assemblers). Hence, the judged importance of the factors was a function of occupational level; content factors were more important for engineers and context for assemblers. Specifically, achievement, work itself, and advancement were most important for the engineers, and job security, company policy and administration, and salary for assemblers.

Correlational analysis showed that the demographic variables cited earlier failed to exert any major influence on these findings regardless of occupational level.

Also, occupationally, the satisfied and dissatisfied subsamples generally followed the same pattern of job factor importance. The

TABLE 5

CORRELATIONS BETWEEN THE JOB FACTOR SATISFACTION SCALES AND THE OVERALL JOB SATISFACTION INDEX FOR THE TWO OCCUPATIONAL SAMPLES AND THEIR SATISFIED AND DISSATISFIED SUBSAMPLES

Factor	Engineers			Assemblers		
	Total ($N = 200$)	Satisfied ($N = 165$)	Dissatisfied ($N = 31$)	Total ($N = 153$)	Satisfied ($N = 101$)	Dissatisfied ($N = 34$)
Recognition	.40**	.22**	.28	.46**	.52**	.28
Responsibility	.59**	.34**	.32	.24**	.10	−.01
Achievement	.76**	.55**	.36*	.64**	.33**	.52**
Advancement	.46**	.38**	−.07	.64**	.56**	.43*
Work itself	.79**	.54**	.52**	.61**	.33**	.39*
Job content score	.76**	.54**	.43*	.69**	.53**	.50**
Salary	.18**	.12	.00	.37**	.36**	.24
Job security	.18**	.12	−.05	.08	.01	.25
Status	.63**	.43**	.35	.59**	.55**	.39*
Supervision—Technical	.46**	.24**	.22	.47**	.46**	.28
Supervision Interpersonal relations	.36**	.19*	.24	.47**	.48**	.35*
Interpersonal relations—Peers	.37**	.09	.18	.31**	.35**	.27
Company policy and administration	.16*	.18*	.01	.38**	.30**	.23
Working conditions	.13	.13	−.10	.29**	.28**	−.08
Job context score	.52**	.31**	.26	.60**	.57**	.44**

* $p \leq .05$.
** $p \leq .01$.

TABLE 6

MEANS, STANDARD DEVIATIONS, AND RANKS OF THE JOB FACTOR IMPORTANCE SCALES SCORES FOR THE ENGINEERS AND THE ASSEMBLERS

Factor	Engineers			Assemblers		
	$\bar{X}$	*SD*	Rank	$\bar{X}$	*SD*	Rank
Recognition	14.3	2.91	10.0	14.5	3.65	11
Responsibility	14.8	2.62	8.0	13.3	3.68	12
Achievement	16.8	2.24	1.0	15.3	3.71	8
Advancement	16.6	2.76	3.0	15.2	4.05	9
Work itself	16.7	2.47	2.0	14.7	3.71	10
Job content score	79.2	9.85		73.1	15.73	
Salary	15.1	3.25	4.5	16.9	3.61	3
Job security	13.9	4.05	11.0	18.2	2.50	1
Status	12.2	3.19	13.0	12.3	4.12	13
Supervision—Technical	15.1	2.72	4.5	16.4	2.95	5
Supervision—Interpersonal relations	15.0	2.36	6.5	16.6	2.74	4
Interpersonal relations—Peers	14.5	2.62	9.0	15.5	3.42	7
Company policy and administration	15.0	2.97	6.5	17.1	2.48	2
Working conditions	13.2	3.00	12.0	15.8	2.97	6
Job context score	114.1	16.05		128.8	17.45	

Note.—$N = 200$ for engineers, $N = 153$ for assemblers.

similarity between the ranks of the specific job factors for these subsamples is shown by a Spearman rank correlation of .94 with the engineers, and .81 for the assemblers.

DISCUSSION

The lack of support found for the content-context dichotomy was generally in line with previous research. On the other hand, content factors clustered together to a higher degree than the context type. One possible explanation for this finding involves the characteristics of the two kinds of variables. Specifically, content aspects stem from the work itself. Thus, it logically follows that when the work itself is challenging and interesting, there are increased opportunities for responsibility, achievement, and recognition. This, in turn, enhances the likelihood of advancement. In contrast, context aspects do not spring from any common denominator, but appear to be independent bedfellows. Hence, enhancement of job security is not necessarily accompanied by increased satisfaction with other context aspects, for example, salary or working conditions. In some instances, the specific factors mean different things to the two diverse groups of workers. Recognition generally clustered together with responsibility to a higher degree with the engineers than assemblers. Conversely, recognition clustered together with salary and company policy and administration to a higher degree for the assemblers than engineers.

Confirmation of the first hypothesis provided some support for both theories. Moreover, the correlations found with the engineers for the specific job factors and overall job satisfaction were strikingly similar to those shown in research which used high-level groups (Halpern, 1966; Wernimont, 1966). Finally, the estimate of overall job satisfaction obtained using the reduced statistical model was also similar to Halpern's results.

For the rejection of the second hypothesis, the findings for the assemblers were generally in line with those of Malinovsky and Barry (1965). Thus, additional evidence was provided for the saliency of the content factors regardless of occupational level. The results also argue against the common stereotype held for semiskilled, blue-collar workers, that is, that they merely trade their time at work and seek security and economic features in order to cope with life's basic needs. While these data appear contradictory to past occupational level research, Centers' (1948) early study also revealed that some of the most liked job factors were of an intrinsic nature for his semiskilled *S*s. In terms of the two theories, doubt was cast on the validity of Darley and Hagenah's (1955) model, and some support provided for Herzberg's theory. Lack of support for Darley's rationale could not be explained as a function of the sample, since these blue-collar assemblers were a poorly educated and moderately paid group of semiskilled workers. Second, it appeared that they had chosen their jobs more as a means for survival or entry into the organization, than for intrinsic factors. Finally, despite the fact that their assignments typically involved repetitive tasks, the majority were relatively satisfied with their jobs.

On the other hand, because advancement is classed as a context factor by Darley and Hagenah (1955), it could not be stated that their theory was completely refuted. This was because in the reduced job factor composites, the relative contributions of the two kinds of specific factors to overall job satisfaction differed as a function of occupational level. Specifically, for the assemblers, the two context factors (as designated by Darley) of advancement and supervision-interpersonal relations, in combination, had a relatively greater correlation with job satisfaction than did achievement, the one content factor.

The theoretical implications were also mixed for the added findings concerned with occupational level differences in the relationships of the variables. Since a number of the relationships differed as a function of occupational level, these findings were contrary to Herzberg's formulations. Furthermore, the results were only partially in line with Darley's rationale (primarily for the content aspects).

The additional findings for the satisfied and dissatisfied subsamples did not fit in with Herzberg's theory. Specifically, the proposed unidirectionality of the two types of factors was not confirmed. That is, job content factors did not operate alone as satisfiers. Instead, both aspects exhibited significant contributions to the criterion in job satisfaction and, hence, both types operated as satisfiers. Finally, the content factors, not the context kind, operated as the most potent dissatisfiers. Moreover, these findings were also contrary to Darley and Hagenah (1955) because occupational level failed to be the main indicator of the kind of factor involved in job satisfaction and dissatisfaction.

Since the judged importance of the two types of factors was found to be a function of occupational level, the predictions of the Darley and Hagenah (1955) theory were supported, but those derived from Herzberg's model were refuted. In the latter theory, it logically follows that content aspects should be judged most important because they are asserted to be the primary contributors to job satisfaction, independent of occupational level. (Content factors are stipulated to act mainly as satisfiers.)

On one hand, this set of results was directly in line with the bulk of research on occupational level (Herzberg, Mausner, Peterson, & Capwell, 1957), and similar to the findings of Centers and Bugental (1966) and Friedlander (1965a).

Conversely, the present findings were contrary to those of Fine and Dickman (1962). Although the discrepancy could be a function of a number of variables, the use of the two-dimensional occupational classification system by Moser, Dubin, and Shelsky (1956), appears to shed some light on the problem. In line with this system, which includes both level and field of work, the lower groups used by Fine and Dickman (electronic technicians and aides, secretaries, and clerks) appear to be of mixed levels and from different fields.

In essence, the implications of the present findings for the two theories differed as a function of the dimension used to rate the job factors (satisfaction versus importance ratings). Neither theory could accommodate both sets of results. From an applied standpoint, the study points up the need to distinguish between occupational level and the dimension used to gauge a job factor. Also, some of the confusion in the research concerned with occupational level and job satisfaction may be traced to the inadequate control of field of work.

Despite the fact that the two theories failed to represent valid models, job satisfaction remains a vital concept for future research. The importance of the concept is noted in

Super's (1957) work where it is viewed as an internal aspect of adjustment to the world of work. Also, inasmuch as a job provides outlets for personality characteristics and basic needs, job satisfaction is conceived as having an important effect upon success and satisfaction in nonjob areas of life (Thompson, 1963).

REFERENCES

ARMSTRONG, T. B. Occupational level as an indicator of satisfiers and dissatisfiers: A test of the Herzberg theory. Unpublished doctoral dissertation, Teachers College, Columbia University, 1969.

BURKE, R. J. Are Herzberg's motivators and hygienes unidimensional? *Journal of Applied Psychology*, 1966, **50**, 317–321.

CENTERS, R. O. Motivational aspects of occupational stratification. *Journal of Social Psychology*, 1948, **28**, 187–217.

CENTERS, R., & BUGENTAL, D. E. Intrinsic and extrinsic job motivations among different segments of the working population. *Journal of Applied Psychology*, 1966, **50**, 193–197.

DARLEY, J. G., & HAGENAH, T. *Vocational interest measurement*. Minneapolis: University of Minnesota Press, 1955.

DEUTSCH & SHEA, INC. *A new look at engineer attitudes*. New York: Industrial Relations News, 1961.

DRAPER, N. R., & SMITH, H. *Applied regression analysis*. New York: Wiley, 1967.

EWEN, R. B. Some determinants of job satisfaction: A study of the generality of Herzberg's theory. *Journal of Applied Psychology*, 1964, **48**, 161–163.

FINE, S. A., & DICKMAN, R. A. Satisfaction and productivity. Paper presented at the meeting of the American Psychological Association, St. Louis, September 1962.

FRIEDLANDER, F. Comparative work value systems. *Personnel Psychology*, 1965, **18**, 1–20. (a)

FRIEDLANDER, F. Relationships between the importance and the satisfaction of various environmental factors. *Journal of Applied Psychology*, 1965, **49**, 160–164. (b)

FROEHLICH, H. P. Factor analysis of an engineering attitude survey. Unpublished masters thesis, Iowa State University, 1958.

HALPERN, G. Relative contributions of motivator and hygiene factors to overall job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 198–200.

HERZBERG, F., MAUSNER, B., PETERSON, R. O., & CAPWELL, D. F. *Job attitudes: Review of research and opinion*. Pittsburgh: Psychological Services of Pittsburgh, 1957.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. *The motivation to work*. New York: Wiley, 1959.

HINRICHS, J. R., & MISCHKIND, L. A. The empirical and theoretical limitations of the two-factor hypothesis of job satisfaction. *Journal of Applied Psychology*, 1967, **51**, 191–200.

HOPPOCK, R. *Job satisfaction*. New York: Harper, 1935.

HOUSE, R. J., & WIGDOR, L. A. Herzberg's dual-factor theory of job satisfaction and motivation: A review of the evidence and a criticism. *Personnel Psychology*, 1967, **20**, 369–389.

HULIN, C. L., & SMITH, P. A. An empirical investigation of two implications of the two-factor theory of job satisfaction. *Journal of Applied Psychology*, 1967, **51**, 396–402.

LINDSAY, C. A., MARKS, E., & GORLOW, L. The Herzberg theory: A critique and reformulation. *Journal of Applied Psychology*, 1967, **51**, 330–339.

MALINOVSKY, M. R., & BARRY, J. R. Determinants of work attitudes. *Journal of Applied Psychology*, 1965, **49**, 446–451.

MOSER, H. P., DUBIN, W., & SHELSKY, I. M. A proposed modification of the Roe occupational classification. *Journal of Counseling Psychology*, 1956, **3**, 27–31.

MYERS, M. S. Who are your motivated workers? *Harvard Business Review*, 1964, **42**, 73–88.

SCHWARTZ, M. M., JENUSAITIS, E., & STARK, H. Motivational factors among supervisors in the utility industry. *Personnel Psychology*, 1963, **16**, 45–53.

SUPER, D. E. *The psychology of careers*. New York: Harper, 1957.

THOMPSON, A. S. Homo laborans—An analysis of the meaning of work. *Bulletin of International Association of Applied Psychology*, 1963, **12**, 5–23.

WALKER, H. M., & LEV, J. *Statistical inference*. New York: Henry Holt, 1953.

WERNIMONT, P. F. Intrinsic and extrinsic factors in job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 41–50.

(Received November 10, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 66–72

EFFECT OF INTERVIEW INFORMATION IN ALTERING VALID IMPRESSIONS[1]

ROBERT E. CARLSON[2]

Life Insurance Agency Management Association,
Hartford, Connecticut

Researchers in the psychology of impression formation have isolated a number of variables which have consistently demonstrated an effect on judges' evaluations. Some authors have extended these findings to the employment setting where interviewers evaluate job applicants based on impressions gained from a fact-finding interview. The feasibility of generalizing from studies where college students evaluate a target *S* as a potential friend, roommate, lover, etc., based on trait adjectives is questioned. The results of a study extending some of the variables of impression formation to a simulated employment interview setting are presented.

The effects of different information properties on assessment evaluations have been both extensively and intensively studied. At this point, for example, is it safe to conclude that negative information has a greater impact on assessment evaluations than does positive information (Bloom & Brundage, 1947; Bolster & Springbett, 1961; Carlson & Mayfield, 1967; Crissy & Regan, 1951; Luchins, 1957a, 1957b; Miller & Rowe, 1967; Newman, Bobbit, & Cameron, 1946; Richey, McClelland, & Shimkunas, 1967; and Springbett, 1958)? Can it also be concluded that information elicited early in the assessment process possesses greater weight (primacy effects) than does information elicited later (recency effects) (Anderson, 1965; Anderson & Barrios, 1961; Asch, 1946; Bolster & Springbett, 1961; and Luchins, 1957a, 1957b)? Is it also true that when order of presentation and type of information are both varied, negative information has a greater and more permanent effect irrespective of its order of presentation (Bolster & Springbett, 1961; Richey, et al., 1967)? In general, the evidence in support of these questions is more consistent in terms of quantity and quality than contrary evidence. Superficially, it is possible to conclude that negative information carries more evaluative weight than does positive information and that primacy effects seem to predominate.

The effect of the range and number of stimulus *S*s on assessment evaluations has also been extensively studied. It has generally been concluded that a stimulus *S* is awarded a more extreme evaluation when preceded by a stimulus *S* of opposite value (Carlson, 1970; Holmes & Berkowitz, 1961; Ohnesorge, 1968; and Rowe, 1967). It has also been shown that judges evaluate a stimulus *S* more conservatively when presented singly than when the same stimulus *S* is presented as a member of a sample (Carlson, 1968).

The effect of this research has been to lead some authors to conclude that evaluations made in applied settings may be a product of fortuitous conditions (Rowe, 1967). That as a result, the employment interview, for example, may not be evaluating an individual, but rather the evaluation may be a product of the situation in which an applicant unwittingly places himself/herself. It should be pointed out, however, that these conclusions, for the most part, have been reached in studies using college sophomores as interviewers and adjectives as interview information. There does appear to be some risk in generalizing to the employment interview from studies where college sophomores evaluate a stimulus *S* as a potential friend or roommate on the basis of four adjectives with such obvious differences as quarrelsome versus capable. Implicit in such generalizations is that the experimental situation closely approximates the employment setting. If the experimental setting contains all the elements of the employment setting, it can be dignified by terming it a model. Obviously, such a goal is laudable, but currently difficult to achieve. However, at the minimum, the experimental setting should attempt to approximate the applied setting on the experimental dimensions in question.

---

[1] This article has been accepted for publication by outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Robert E. Carlson, Director of Developmental Research, Life Insurance Agency Management Association, 170 Sigourney Street, Hartford, Connecticut 06105.

The research reported here is an attempt to extend to the employment setting some of the variables which have consistently demonstrated an influence on assessment evaluations in less applied circumstances. These variables are the favorability of the information, the order in which the information is presented within a stimulus *S*, the order in which the stimulus *S*s are evaluated, and the validity of the information being evaluated. In addition, the judges were first trained on use of valid information, the judges all had conducted interviews and the administration of the experimental material followed as closely as possible the sequence of a selection system. The purpose of the study was to determine to what extent an initial valid impression is altered by selection interview information differing in favorability, order of presentation, and quality of the preceding applicant.

METHOD

Subjects

The responding *S*s who acted as interviewers were life insurance agency managers who were attending 2-wk. Schools in Agency Management conducted by the Life Insurance Agency Management Association. As managers, they have complete responsibility for recruiting, interviewing, selecting, and training salesmen to work under their supervision.

The school hours cover all aspects of the manager's job. Specifically, selection hours deal with the value, content, and conduct of the selection interview, tests, and early training. The managers are well indoctrinated on the validity and use of the Aptitude Index Battery (AIB; Paterson & Thompson, 1953; Peterson & Wallace, 1966). The success base rates for various scores are presented and managerial costs and implications are discussed. The managers are further instructed that as a result of the triangular bivariate distribution between test score and the criteria, the test is more accurate in predicting failure than success. Thus, the managers are subject to a highly standardized and practical interpretation of test results on the AIB.

Construction of Hypothetical Job Applicants

Previous research has clearly indicated that meaningful results can be obtained by constructing hypothetical job applicants from prescaled descriptive statements (Webster, 1964). The type of statement used obviously depends on the situation to which *E* wishes to generalize his results. In the research reported here, the goal was to simulate the fact-finding interview. The type of statements used were descriptive personal history items. The development of the personal history information used to construct the hypothetical applicants has been discussed elsewhere (Carlson, 1967; Mayfield & Carlson, 1966).

The life insurance agency managers attending the 2-wk. schools each evaluated eight hypothetical job applicants during the second week of the school. The first item of information about an applicant was always the aptitude test results. For one-third of the sample the test results indicated that the experimental applicant had passed the AIB, for one-third of the sample the hypothetical applicant had not taken the AIB, and for one-third of the sample the applicant had failed the Aptitude Index test. The test results were initially presented alone on the top of the first page of the test booklet describing an applicant and repeated as the first item on all succeeding pages describing an applicant. The remaining eight items of personal history information were presented one at a time in a serial-cumulative fashion. Each test booklet thus consisted of nine pages with each page adding an additional item of information. At the bottom of each page the manager was asked to rate the applicant on a 7-point scale as to his potential success as a life insurance sales representative. Thus, each manager was asked to rate each hypothetical applicant nine different times based on nine different sets of information.

Four different types of hypothetical applicants served as the stimulus material. The first type of hypothetical applicant was constructed from the test results and eight items of favorable information (FF). The items were approximately equal in scale value, standard deviation, and test-retest reliability. A second type of applicant was constructed from the test results and eight items of unfavorable information (UU), again with approximately the same mean, standard deviation, and test-retest reliability. A third hypothetical applicant was constructed from the test results and four items of favorable information and four items of unfavorable information (FU). The four items of favorable information were the same as the first four items of favorable information describing the first applicant, while the four items of unfavorable information were the same as the first four items describing the second applicant. The fourth applicant was the same as the third, except the four items of unfavorable information were presented first and the four items of favorable information presented last (UF).

For all hypothetical applicants, only the scale value varied, the standard deviation and test-retest reliability of the items were approximately equal. One-fourth of the sample responded to each of the four types of hypothetical applicants.

In one-half of the test booklets the hypothetical applicant preceding the experimental applicant was of similar stimulus value, but constructed of different items of personal history information. In the remaining half of the test booklets the preceding job applicant was of opposite value. The experimental pair was randomly allocated among the remaining six applicants.

Analysis

The original design of the experimental material was a 2 × 3 × 4 factorial with nine repeated ratings made by each interviewer. Because the items of personal history information were similar in stimulus value in blocks of four, it was decided prior to analysis to analyze three levels or amounts of information rather than nine. The first level of information was the test results alone (T), a second amount of information was the test results plus the first four items of personal history information (T + 4), while the third level of information was the test results plus all eight items of personal history information (T + 8). In addition, rather than encounter unequal variance-covariance matrices as a result of the repeated measures and also to allow additional treatment effects to be explored, a random sample of 25 *S*s was taken within each cell. The resultant design was a 2 × 3 × 4 × 3 factorial with 25 observations per cell. The independent variables were two types of applicant samples (homogeneous, where the experimental and preceding job applicants were of similar stimulus value; heterogeneous, where the experimental and preceding job applicants were of opposite value), three levels of test results (pass, not taken, failed), four types of personal history information (FF, FU, UF, UU), and three amounts of information (T, T + 4, T + 8). A factorial analysis of variance was performed on the 2 × 3 × 4 × 3 treatment combinations.

RESULTS

Table 1 presents the mean and standard deviation of the ratings for the treatment combinations.

Table 2 presents the results of the analysis of variance applied to the interviewer's evaluations. Looking first at the *F* values and their statistical significance, it could superficially be concluded that almost all the independent variables and their various combinations produced some effect on the interviewer's evaluations. This would be especially true if concentration were focused on the number of significant *F* values and their magnitude. However, Table 2 also presents omega-squared (Hays, 1963) which indicates the relative effect of the treatments on the dependent variable. By first determining whether or not a particular treatment produces a statistically

TABLE 1

MEANS AND STANDARD DEVIATIONS OF INTERVIEWER'S EVALUATIONS

Test results	Homogeneous sample						Heterogeneous sample					
	T[a]		T + 4[b]		T + 8[c]		T		T + 4		T + 8	
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
Passed test												
FF	3.6	.6	3.9	.8	4.2	.9	3.5	1.1	4.2	1.1	5.6	.9
FU	3.7	.7	4.0	.9	1.2	1.0	3.4	.5	4.3	.6	2.4	1.2
UF	3.6	.7	.8	.9	1.6	.9	3.4	.5	2.2	1.1	3.1	1.2
UU	3.7	.8	1.1	.9	1.1	1.1	3.5	.7	1.0	1.1	.8	1.3
Not-taken test												
FF	3.1	.6	3.5	.9	3.6	1.0	3.0	.8	4.1	.8	4.8	.9
FU	3.2	1.0	4.2	1.0	1.2	1.1	2.7	.7	4.0	.9	2.1	.9
UF	3.1	.6	1.1	.9	1.5	1.1	2.9	.6	1.8	.8	3.0	1.2
UU	3.0	1.1	1.0	.9	.8	1.1	2.9	.5	.8	.9	.7	1.1
Failed test												
FF	1.2	.9	1.8	1.0	2.2	1.0	1.2	.9	2.1	1.2	2.8	1.2
FU	1.3	.9	2.2	1.0	.6	.9	1.0	.9	2.2	1.2	.9	.9
UF	1.2	1.0	.4	.7	.5	.9	1.0	.8	.7	.7	1.9	1.1
UU	1.2	.9	.4	.6	.4	.8	1.1	.9	.3	.7	.2	.7

Note.—*N* = 25 in each cell. Abbreviations: FF = favorable information, UU = unfavorable information, FU = favorable and unfavorable information, and UF = unfavorable and favorable information (see text).

[a] T = test results alone.

[b] T + 4 = test results plus four items of personal information.

[c] T + 8 = test results plus eight items of personal information.

significant effect on the interviewer's evaluations, and second, the magnitude of this effect, the importance of the treatment effects are more realistically evaluated.

Application of both the statistical and practical criteria point out three treatment conditions which appear to have an effect on an interviewer's evaluation of a hypothetical applicant. First, the most important, 14% of the explainable variance in the interviewer's evaluations is attributable to the objective information, valid test results. Second, another 12% of the variance in interviewer's evaluations is due to differences in the favorability of subjective personal history information. Third, an additional 11% of the variance in the interviewer's evaluations is attributable to a combination of differing amounts and types of information. The remaining treatment effects, while statistically significant, do not practically reduce the variability of interviewer's evaluations.

TABLE 2

ANALYSIS OF VARIANCE OF INTERVIEWER'S RATING

Source	*df*	*MS*	*F*	W^2
Amount of information (I)	2	49.65	42.08*	.01
Type of information (J)	3	289.54	245.37*	.12
Sample (K)	1	27.89	23.64*	.00
Test results (L)	2	497.59	421.69*	.14
IJ	6	127.41	107.97*	.11
IK	2	32.54	27.58*	.01
IL	4	19.75	16.74*	.01
JK	3	19.20	16.27*	.01
JL	6	9.05	7.67*	.01
KL	2	.79	.67	.00
IJK	6	6.09	5.16*	.00
IJL	12	2.96	2.51*	.00
IKL	4	.89	.75	.00
JKL	6	1.43	1.21	.00
IJKL	12	.75	.64	.00
Error	1728	1.18		

* $p = .01$.

Figure 1 presents a diagram of the interviewer's mean evaluations of the hypothetical applicants. The interaction of the amount of information with the type of information is presented for the three different test-result conditions. By visual examination, it is clear the interaction is built into the experimental setting by the type of stimulus applicant constructed. However, the interaction per se is not the major interest. The primary interest is in the amount of change in initial valid impressions created by different kinds and order of information. Based upon these results, three post hoc hypotheses were developed and tested for statistical significance. The post hoc hypotheses were as follows:

1. For hypothetical applicants constructed of consistent information (all favorable or all unfavorable): (*a*) The first four items of favorable information have the same effect on interviewers' evaluations as the second set of four favorable items $(\bar{X}_{F_T} - \bar{X}_{F_{T+4}} = \bar{X}_{F_{T+4}} - \bar{X}_{F_{T+8}})$. There is neither a primacy or recency effect for favorable information. (*b*) The first four items of unfavorable information have a significantly greater effect on interviewers'

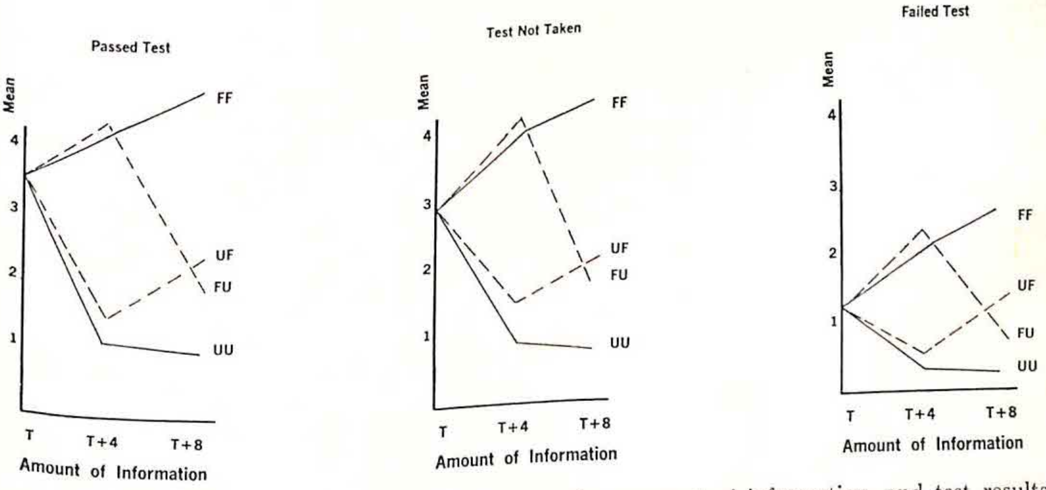


FIG. 1. Interviewer's mean rating by type of information, amount of information, and test results.

evaluations than the second set of four unfavorable items ($\bar{X}_{U_T} - \bar{X}_{U_{T+4}} > \bar{X}_{U_{T+4}} - \bar{X}_{U_{T+8}}$). There is a primacy effect for unfavorable information.

2. For hypothetical applicants constructed of inconsistent information (half favorable, half unfavorable): (*a*) Unfavorable information produces greater change in an inverviewer's evaluation when presented last, that is, a recency effect ($\bar{X}_{FU_T} - \bar{X}_{FU_{T+8}} > \bar{X}_{UF_T} - \bar{X}_{UF_{T+8}}$).

These three hypotheses were tested by the method of nonorthogonal contrasts. The three hypotheses were accepted at the .01 probability level. Thus, the effect of subjective interview information in altering valid impressions is dependent on the amount, kind, and order in which it is presented. However, it should be recalled that while valid impressions may be altered, they still account for the largest single portion of the explainable variance in the interviewer's evaluations.

DISCUSSION

The substantive findings from the original analysis indicated that three variables accounted for significant portions of the variance in the interviewer's evaluations. The first independent variable, valid test information, produced the greatest effect on interviewer's evaluations. This confirms a previous study in which the author found that without subjective personal information, the test results accounted for 55% of the explainable variance in the interviewer's evaluation (Carlson, 1970). In the research reported here, however, the test results accounted for only 14% of the criterion variance. The obvious inference is that the presence of the subjective information and the contextual variables, through their interaction with the test information, have diminished the absolute effect of the valid test information. To examine this explanation, Table 1 was collapsed to two dimensions (test results and quality of previous applicant) by eliminating the effects of different types of subjective information (T + 4, T + 8; and FF, FU, UF, and UU) and by performing a 3 × 2 analysis of variance. This analysis indicated that the test results accounted for 30% of the explainable variance. Thus, the difference between the 55% of the variance accounted for in the previous study and the 30% accounted for in this study is due to sample and design differences, while the difference between the 30% of the variance accounted for in the 3 × 2 analysis and the 14% of the variance accounted for in the 3 × 2 × 4 × 3 original analysis is due to the interaction of the test results with the additional independent variables. Thus, initial valid impressions can be readily altered by subjective personal information.

The second independent variable, favorability of the information used to construct a candidate, accounted for 12% of the explainable variance in the interviewer's evaluations. As would be expected, the applicants constructed of eight items of favorable information were rated most positively while the applicants constructed of eight items of unfavorable information were rated most negatively. Of the applicants constructed of half favorable–half unfavorable information, the UF order produced more positive ratings than the FU order. This persisted irrespective of the type of test results used to establish the original valid impression.

The interaction between type of information and amount of information, which accounted for another 11% of the explainable variance, indicates to some extent the mechanism by which the initial valid impressions are altered. The interviewers seem to be demonstrating the rule that "when the selection test is failed, forget the man"; and "when the selection test is passed, forget the test and evaluate the applicant using other tools." As would be expected, information of opposite value to the initial valid impressions produced a much greater change in rating than did information of similar value. The initial evaluation of a hypothetical job applicant who has either passed or not taken a valid selection test is altered to a greater extent by unfavorable information and much less affected by favorable information. The initial evaluation of a hypothetical job applicant who has failed a valid selection test is only slightly more affected by favorable than by unfavorable information. In general, information which is consistent with an interviewer's initial impression adds less to his total impression than information of opposite value detracts from it. It thus appears interviewers may have some difficulty in assimilating and interpreting information of similar value, while information of obviously different values can be easily interpreted and possibly overreacted to. The extent to which an interviewer feels uncomfortable with inconsistent information may serve to explain why Webster (1964) found that interviewers actively sought information to support their early impressions.

The results of the tests on the post hoc hypotheses again demonstrate the strong effect of unfavorable information and suggest further that both primacy and recency effects may be created by unfavorable subjective information. Unfavorable information produced primacy effects when the hypothetical job applicants were constructed of uniform or consistent information. In this situation the first set of unfavorable information accounted for the majority of the change in interviewer's evaluations. Additional unfavorable information changed the initial negative evaluation very little. This was observed following all three test-result situations. This indicates that the contrast effect of negative information with any preestablished set may generate a primacy effect. Thus, only a few items of unfavorable information are necessary to create a negative impression; adding more unfavorable information does not make the evaluation any more negative. By comparison, favorable information showed neither a primacy nor a recency effect. Favorable impressions are created much less rapidly and increases in positive evaluations do occur with additional information.

The results for applicants of contrasting subjective personal history information suggest a recency effect for unfavorable information. There is a greater change in the initial impression created by valid information when subjective information is presented in a favorable-unfavorable order than when the same information is presented in an unfavorable-favorable order. Although the information is exactly the same, the effect of unfavorable information is greater after a positive impression has been established than is the effect of positive information after a negative impression is created. Thus, while both recency and primacy effects can be created by unfavorable information (after both positive and negative initial valid sets have been established), it is difficult to conclude that this demonstrates anything beyond the strength of unfavorable subjective information in creating a contrasting situation.

While the assimilation of later information into early evaluations makes conceptual sense, the data presented here demonstrated no such effect for favorable information and suggest that unfavorable information produces a contrast effect, given either a positive or negative initial impression. It is recognized, however, that the stimulus material employed in the research reported here may have been too obvious (e.g., a college graduate vs. did not graduate from high school) and that more subtle trait descriptions (e.g., honest vs. dishonest) do in fact produce primacy effects which may be due to assimilation of later information into early impressions.

The findings of the research reported here also serve to reinforce the validity of two other current issues in the conduct of laboratory experiments. First, they provide ample evidence in support of Cohen's (1968) statement that ANOVA tables should include a measure of the proportion of variance accounted for by a treatment effect. In this research, two variables shown to have an effect in studies of impression formation generated statistically significant results, but did not practically reduce the variance in the interviewer's judgments. Second, these results indicate that a great deal of caution is necessary in extending the conclusions of laboratory experiments of theoretical interest to applied settings where the goal may be similar, but the task and the tools used to perform the task are different. The laboratory setting or simulation experiment may appear to be very real or natural in its conception, but it does not eliminate the need for testing these findings and hypotheses in natural settings (Barker, 1968; Garnsey & Hibbs, 1967; McGuire, 1967; and Williams & Rausch, 1969).

REFERENCES

ANDERSON, N. H. Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, 1965, 2, 1–9.

ANDERSON, N. H., & BARRIOS, A. A. Primacy effects in personality impression formation. *Journal of Abnormal and Social Psychology,* 1961, **63**, 346–350.

ASCH, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology,* 1946, **41**, 258–290.

BARKER, R. G. *Ecological psychology*. Stanford: Stanford University Press, 1968.

BLOOM, B. F., & BRUNDAGE, R. G. Prediction of success in elementary schools for enlisted personnel. In P. M. Street (Ed.), *Personnel research and test development*. Princeton: Princeton University Press, 1947.

BOLSTER, B. I., & SPRINGBETT, B. M. The reaction of interviewers to favorable and unfavorable information. *Journal of Applied Psychology,* 1961, **45**, 97–103.

CARLSON, R. E. Selection interview decision: The effect of interviewer experience, relative quota situation, and applicant sample on interviewer decisions. *Personnel Psychology,* 1967, **20**, 259–280.

CARLSON, R. E. Selection interview decisions: The effect of mode of applicant presentation on some outcome measures. *Personnel Psychology,* 1968, **21**, 193–207.

CARLSON, R. E. Effect of applicant sample on ratings of valid information in an employment setting. *Journal of Applied Psychology,* 1970, **54**, 217–222.

CARLSON, R. E., & MAYFIELD, E. C. Evaluating interview and employment application data. *Personnel Psychology,* 1967, **20**, 441–460.

COHEN, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 1968, **70**, 426–443.

CRISSY, W. J., & REGAN, J. Halo in the employment interview. *Journal of Applied Psychology,* 1951, **35**, 338–341.

GARNSEY, M. E., & HIBBS, J. R. *Social sciences and the environment*. Boulder: University of Colorado Press, 1967.

HAYS, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

HOLMES, D. S., & BERKOWITZ, L. Some contrast effects in social perception. *Journal of Abnormal and Social Psychology,* 1961, **62**, 150–152.

LUCHINS, A. S. Experimental attempts to minimize the impact of first impressions. In C. I. Hovland (Ed.), *The order of presentation in persuasion.* New Haven: Yale University Press, 1957, 62–75. (a)

LUCHINS, A. S. Primacy-recency in impression formation. In C. I. Hovland (Ed.), *The order of presentation in persuasion*. New Haven: Yale University Press, 1957, 33–61. (b)

MAYFIELD, E. C., & CARLSON, R. E. Selection interview decisions: First results from a long-term research project. *Personnel Psychology,* 1966, **18**, 41–55.

MCGUIRE, W. Some impending reorientations in social psychology: Some thoughts provoked by Kenneth Ring. *Journal of Experimental and Social Psychology,* 1967, **3**, 124–139.

MILLER, J. W., & ROWE, P. M. The influence of favorable and unfavorable information upon assessment decisions. *Journal of Applied Psychology,* 1967, **51**, 432–435.

NEWMAN, J., BOBBIT, J. M., & CAMERON, D. C. The reliability of the interview method in officer candidate evaluation program. *American Psychologist,* 1946, **1**, 103–109.

OHNESORGE, J. P. Interviewer evaluation of job applicants as a function of qualifications of the immediately preceding applicants: A demonstration of contrast effects. Paper presented at Midwestern Psychological Association Convention, May 1968.

PATERSON, D. G., & THOMPSON, A. S. Review of aptitude index. In Oscar K. Buros (Ed.), *The fourth mental measurements yearbook*. New Jersey: Gryphon Press, 1953.

PATERSON, S. A., & WALLACE, S. R. Validation and revision of a test in use. *Journal of Applied Psychology,* 1966, **50**, 13–17.

RICHEY, M. H., MCCLELLAND, L., & SHIMKUNAS, A. M. Relative influence of positive and negative information in impression formation and persistency. *Journal of Personality and Social Psychology,* 1967, **6**, 322–327.

ROWE, P. N. Order effects in assessment decisions. *Journal of Applied Psychology,* 1967, **51**, 170–173.

SPRINGBETT, B. M. Factors affecting the final decision in the employment interview. *Canadian Journal of Psychology,* 1958, **12**, 13–22.

WEBSTER, E. C. *Decision-making in the employment interview*. Montreal: Industrial Relations Centre, McGill University, 1964.

WILLIAMS, E. P., & RAUSCH, H. L. *Naturalistic viewpoints in psychological research*. New York: Holt, Rinehart and Winston, 1969.

(Received November 12, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 73–79

RELEVANCY IN THE TEST OF MOTIVATOR-HYGIENE THEORY

BENEDICT S. GRIGALIUNAS[1] AND FREDERICK HERZBERG

Case Western Reserve University

Motivator-hygiene (M-H) theory predictions have been replicated through the sequence of events methodology, but when alternative methodologies are utilized, inconsistent results are often obtained. Item rating and rating procedures have been accused of producing irrelevant data for the study of job attitudes. Therefore, a study comparing the two methodologies was made. A measure of relevance was obtained in order to determine its effect in producing the empirical inconsistencies. The results showed that ratings and rankings of items measuring job attitudes are heavily contaminated by irrelevancy. In addition to irrelevancy, the failure of rating and ranking of items to replicate M-H theory predictions seems to be further due to the biases produced by social desirability, value systems, and misinterpretations or reinterpretation of items by the respondents.

The fundamental prediction of the motivator-hygiene (M-H) theory of Frederick Herzberg, initially reported in the *Motivation to Work* (Herzberg, Mausner, & Snyderman, 1959), has undergone well over 50 replications and is, perhaps, the most replicated study in contemporary industrial psychology. The sequence of events methodology and its variations has been largely utilized (though not exclusively) throughout these replications, and consequently, this has led to the criticism that the M-H theory is "shackled to the story telling methodology [Dunnette, Cambell, & Hakel, 1967, p. 148]." Also, there exists evidence of inconsistency in results where alternative methodologies are utilized.

Herzberg et al. (1959) chose the sequence of events methodology in their original study for three reasons: (*a*) Attitudinal research is not yet sure of what an attitude is and how it may differ from similar concepts such as opinions, interests, feelings and tendencies; (*b*) How can one know that someone has an attitude or feeling about something? Employees are often given predetermined choices of job factors to rate or check, and as a result, attitudes are given when none may exist; (*c*) Cultural noise responses are too often the product of rating scale items, particularly when they are irrelevant to the job situation (Herzberg et al., 1959).

Because of the above problems in attitudinal research, the criteria for a job attitude sequence were defined as follows: it had to be an objective event, something had to have happened; it had to be bounded by time (beginning, middle, and end); the feelings had to be exceptionally good or bad; and the respondents' feelings about his job had to be directly affected. As a result, the focus of the data, because the respondent was personally involved, was on the existence and the change of a feeling. The criteria ensured that a feeling was being tapped (Herzberg, 1966). In this manner, factors emerge out of the data rather than being determined a priori.

Herzberg's methodological rationale has led Dunnette and Kirchner (1965) to state:

> A more fundamental contribution of the study is that job factors so identified were allowed to emerge from descriptions of actual job situations rather than being based exclusively on responses to checklists or sets of statements developed ahead of time by the investigator. The job factors derived by Herzberg's classification are more likely, therefore, to reflect things in the job environment leading to employees' approach and avoidance behaviors [p. 152].

Yet, many if not most of the research studies whose results are inconsistent with the M-H studies have utilized rating and ranking procedures of factors or items. A few, for example, are: Burke's (1966) ranking of 10 job characteristics in order of importance for the self; Dunnette et al.'s (1967) Q-sort decks for satisfying and dissatisfying job situations;

---

[1] Requests for reprints should be sent to Benedict S. Grigaliunas, Department of Psychology, Case Western Reserve University, Cleveland, Ohio 44106.

Ewen, Smith, and Hulin's (1966) job descriptive index with GM faces scale; Friedlander's (1965) questionnaire based on 14 work environment characteristics and their importance, and Friedlander's (1964) ratings of importance of 18 factors; Graen's (1966) ratings or rankings of importance of Herzberg's factors; Rosen's (1963) rating of importance of absence of items; and Wernimont's (1966) forced-choice and free-choice rating of Herzberg's factors.

These methodologies have been accused of producing irrelevant data by Whitsett and Winslow (1967). Especially cited is the Wernimont (1966) and the Dunnette (1965) studies in which the problem of irrelevant rating of items is most apparent. Wernimont obtained responses to both forced-choice and free-choice items from 50 accountants and 82 engineers in self-descriptions of past satisfying and dissatisfying job situations. He constructed two forced-choice checklists, each made up of 50 pairs of items with each pair consisting of one motivator and one hygiene item. The pairs of items were equated for social desirability and one set of 50 was written in a positive tone and the other in a negative tone. Wernimont (1964) then states:

> Since each respondent was obligated to choose one statement from each of fifty pairs of statements by placing a check mark in front of that statement, it can readily be seen that some of the items chosen might have little or nothing to do with a given specific situation. In other words, the individual would be forced to choose between two statements neither of which would be relevant to the situation he had just described [p. 41].

Nevertheless, Wernimont (1966) concluded that since both motivator and hygiene items had been checked with respect to both happy and unhappy situations, both sets of factors contribute to satisfaction and dissatisfaction.

In the second part of the study, Wernimont (1966) permitted the respondents to freely choose any 10 items which best described the satisfying (dissatisfying) situation on the job. The results then tended in the direction similar to those of Herzberg's studies and the numerous replications. However, Wernimont chose to attribute the M-H result to "respondent defensiveness bias" and accepts the forced-choice result because of its elimination of this bias. In the light of his own words, the opposite is a more logical argument since the irrelevancy is the bias as he points out so well in his discussion of his use of the forced-choice methodology.

In order to examine the irrelevance phenomenon, Stern (1968) used 20 statements taken from a study by Dunnette (1965) with high factor loadings on 10 factors. He had each *S* write both a positive and a negative incident and then rate each statement on a scale of 1–5 of increasing importance of that statement in producing the feeling of satisfaction (dissatisfaction). However, he reserved a rating of 0 to those items which *S* felt to be totally irrelevant.

As a result, Stern (1968) found that *S*s themselves rated approximately 50% of well over 9,000 possible ratings as being totally irrelevant to the high or low incidents described. The remaining 50% of the ratings had 7 or more factors rated from 1 to 5, and 10% of the total ratings had all 10 factors rated as relevant even though his content analysts found that 1 or 2 and rarely 3 or 4 factors could be relevant to any experience. He also found that *S*s rate more motivator factors as relevant in both the happy and unhappy sequences.

The purpose of this study was to test whether irrelevancy is a crucial variable in determining the inconsistent results between M-H theory methodology and rating scale methodologies by analyzing one set of data on the same *S*s by both methodologies. Also, a measure of relevancy or irrelevancy was obtained in order to estimate its effect on the results of the data. In this manner, Herzberg's third reason for utilizing the sequence of events methodology, that cultural noise responses are too often the product of rating scale items when they are irrelevant to the job situation, can be tested.

METHOD

Hypotheses

Hypothesis 1. The content analysis coding will yield the predicted M-H theory result—the high incidents will consist of significantly more motivator sequences while the low incidents will consist of significantly more hygiene sequences.

Hypothesis 2. The *S*s' ratings of the statements will not yield the predicted M-H theory result—both motivator and hygiene statements, but especially motivator statements will be rated highly for both high and low incidents.

Hypothesis 3. When only the relevant ratings are used, the predicted M-H theory outcome will result —the high incidents will consist of significantly more motivator sequences while the low incidents will consist of significantly more hygiene sequences.

Procedure

A total of 81 *Ss*, junior and senior students at Case Western Reserve University, were requested to participate in the study. Their mean age was 21 yr. Forty-three were male and 38 were female students. The College Experience Questionnaire (CEQ) developed by Stern (1968) was utilized to gather the data. Briefly, the CEQ consists of two parts: Part 1 follows the outline of the Herzberg et al. (1959) interview and elicits two incidents, one high and one low, the only difference is that the CEQ incidents are in the briefer written protocol form rather than the richer, more lengthy incidents of the interview. Part 2 further examines the incidents in terms of the reasons given by the respondents for their feelings. The *Ss* are asked to rate 20 statements per experience as to the importance of that statement in producing the high or low feeling for that experience.

The statements include five motivator and five hygiene factors, each factor represented by two statements. The items used had high factor loadings on M-H factors as developed by Dunnette et al. (1967), but were modified in order to better fit a college population.[2] The factors were:

Motivator	Hygiene
Achievement	Co-workers (Peers)
Responsibility	Supervision
Work Itself	Status
Recognition for Achievement	Salary
Personal Growth	Company Policy and Administration

Two sets of statements were included, one set to be used for happy (high) experience and one set for unhappy (low) experiences. Generally, the two sets are written as opposites of each other and the order of statements differs for each set. Each of the statements was rated on a scale from 1 to 5 of increasing importance of that factor in producing the feeling of satisfaction (dissatisfaction).

In order to obtain a measure of relevancy, *Ss*' high and low incidents and the two sets of statements without *Ss*' ratings were given to six judges. Each judge was instructed to read each incident and its corresponding statements and then only check those statements that are relevant to the incident described by *S*. Of the six judges, three were advanced graduate students who were well acquainted with the M-H theory (the expert judges) and three were beginning graduate students who were totally naïve of M-H theory (the naïve judges). This combination of raters was used in order to ensure that there would be no source of bias resulting from a rater's familiarity with the M-H theory.

TABLE 1

MOTIVATOR, HYGIENE, AND MIXED SEQUENCES IN HIGH AND LOW FEELINGS FROM CONTENT ANALYSIS

Sequence	High feelings	Low feelings	Total
Motivator	45	14	59
Hygiene	18	53	71
Mixed	8	10	18
Total	71	77	148

Note.—$\chi^2 = 33.51, p < .001$.

Two different expert judges were then given only *Ss*' high and low incidents and were instructed to content analyze the incidents according to the Herzberg methodology. The coding scheme was as follows: motivator sequence, hygiene sequence, mixed (both motivator and hygiene content) sequence, and uncodable.

RESULTS

Hypothesis 1

Of the 81 *Ss* requested to relate both a high and a low incident, 76 high and 81 low incidents were obtained.[3] The two content analysis coders were in perfect agreement on 83% of their coding. The differences between the two coders were then resolved and 100% agreement was reached. Five high and four low incidents were categorized as uncodable because the incidents did not meet the aforementioned criteria for an incident.

Table 1 presents the coding of the high and low incidents and Figure 1 presents the same data in percentages according to the manner of presentation in Herzberg (1966). Since the mixed category is almost equally divided between the high and low sequences and contributes relatively little to the chi-square, it was excluded from Figure 1A.

It is evident from the data that the basic M-H theory result has been replicated. Of the high incidents, 71% consisted of motivator sequences and 29% consisted of hygiene sequences while the low incidents consisted of

---

[2] Status and personal growth were not included in the Dunnette et al. (1967) study.

[3] Five *Ss* related only low incidents and no high incidents.

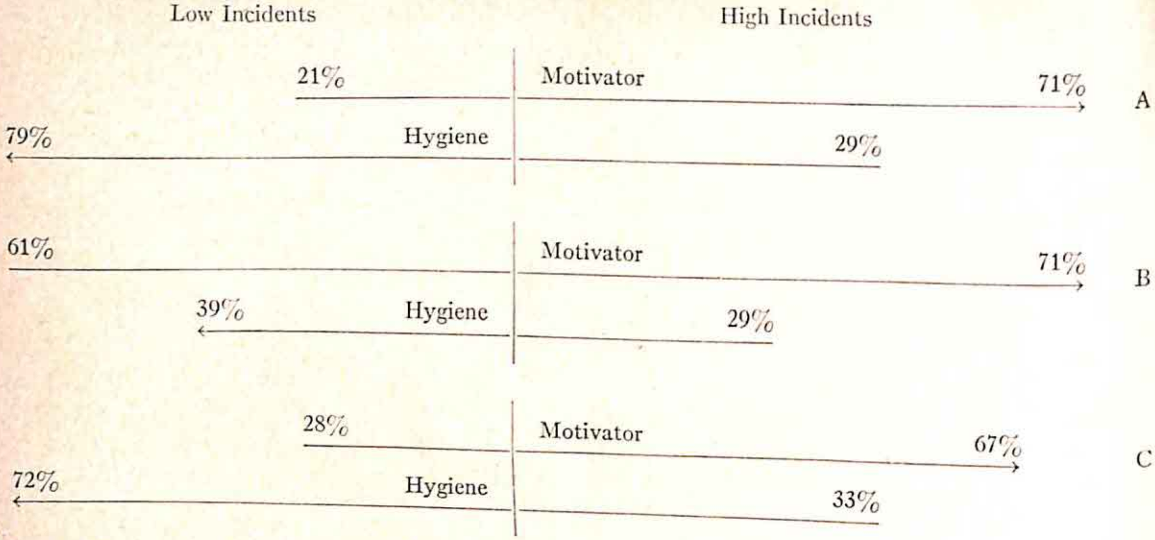


FIG. 1. Percentages of motivator and hygiene sequences in high and low feelings from (A) content analysis, (B) *Ss*' 4 and 5 ratings, and (C) judges relevancy ratings.

79% hygiene sequences and 21% motivator sequences. The chi-square is 33.51, significant well beyond the .001 level.

Hypothesis 2

The *Ss* rated the corresponding sets of statements for 72 high and 75 low incidents.[4] Each statement was rated on a scale from 1 to 5 of increasing importance of that statement in producing the high or low feeling. Although the ratings of 2 and 3 can be considered as indicating some importance, only the ratings of 4 (very important) and 5 (extremely important) were used in the forthcoming analysis. If irrelevant rating does occur and has a significant effect on the data, it would manifest itself in these extreme ratings.

The items rated 4 and 5 by *Ss* were used to categorize the incidents into the following types of sequences: motivator sequence, hygiene sequence, mixed sequence, and uncodable.[5] That is, if all the items rated 4 and 5 by *Ss* were motivator items, then the sequence was categorized as a motivator sequence. Similarly, if the items rated 4 and 5 included both motivator and hygiene items, then the sequence was categorized as a mixed sequence. The results are shown in Table 2 and indicate that an M-H theory result does not occur. There is no significant difference in the number of motivator and hygiene sequences between high and low incidents.

Since *Ss*' ratings produced such a large proportion of mixed sequences, a sequence categorization could no longer be used to examine the percentage of motivator and hygiene sequences in high and low incidents. Too much data is lost if the mixed sequences are excluded. However, the sequences do contain multiple factors, and consequently, the specific factors within all the sequences are grouped in Figure 1B according to the manner of presentation in Herzberg (1966). Again, an M-H theory result is not obtained.

TABLE 2

MOTIVATOR, HYGIENE, AND MIXED SEQUENCES IN HIGH AND LOW FEELINGS FROM *Ss*' 4 AND 5 RATINGS

Sequence	High feelings	Low feelings	Total
Motivator	23	21	44
Hygiene	5	8	13
Mixed	41	41	82
Total	69	70	139

Note.—$\chi^2 = .82$, *ns.*

---

[4] Four high and six low incidents did not have their corresponding statements rated by *Ss*.

[5] Three high and five low incidents made up the uncodable category because they had no statements rated 4 or 5.

Table 3 presents the frequencies of the 4 and 5 ratings in a chi-square table which yields a chi-square of 12.77, significant beyond the .001 level. This data, as well as Table 2 and Figure 1B, indicate that motivator factors obtained through ratings are more important in producing both high and low feelings. Also, both motivator and hygiene factors are seen to be important in contributing to satisfaction and dissatisfaction as claimed by the "traditional" theory of job satisfaction.

Hypothesis 3

All 76 high and 81 low sequences were read by the judges and their corresponding sets of statements were rated on relevancy. A criterion of 4/6 or better agreement of judges' ratings was set in order to determine whether a particular statement was relevant to the incident. Both high and low incidents were then coded according to the relevant statements in the following categories: motivator sequence, hygiene sequence, mixed sequence, or uncodable.[6] That is, if all the items rated relevant by the judges were motivator items, then the sequence was categorized as a motivator sequence. Similarly, if the items rated relevant included both motivator and hygiene items, then the sequence was categorized as a mixed sequence. The preliminary analysis indicated that there were no significant differences in the number or kind of factors rated relevant by the expert and the naïve judges.

TABLE 3

FREQUENCY OF MOTIVATOR AND HYGIENE FACTORS IN HIGH AND LOW FEELINGS FROM *Ss*' 4 AND 5 RATINGS

Factor	High feelings	Low feelings	Total
Motivator	415	246	661
Hygiene	167	160	327
Total	582	406	988

Note.—$\chi^2 = 12.77$, $p < .001$.

TABLE 4

MOTIVATOR, HYGIENE, AND MIXED SEQUENCES IN HIGH AND LOW FEELINGS FROM JUDGES' RELEVANCY RATINGS

Sequence	High feelings	Low feelings	Total
Motivator	41	16	57
Hygiene	20	42	62
Mixed	12	15	27
Total	73	73	146

Note.—$\chi^2 = 19.10$, $p < .001$.

The results are presented in Table 4 and Figure 1C. High incidents consist of 67% motivator sequences and 33% hygiene sequences, while low incidents consist of 72% hygiene sequences and 28% motivator sequences. The obtained chi-square of 19.10 is significant beyond the .001 level. Thus, when only the relevant statements are analyzed, the results are very similar to those obtained by Herzberg (1966) and the replications. This is clearly illustrated by comparing the results of Hypothesis 1 with results of Hypothesis 3.

The mean number of statements rated 4 and 5 by *Ss* and the mean number of statements rated relevant by the judges are presented in Table 5. The *Ss* chose a ratio of approximately three statements to one chosen by the judges. This discrepancy is due to *Ss* giving more of the high ratings to motivator factors as indicated in Table 3.

A good example of the irrelevancy of the ratings given by *Ss* can be illustrated by the following case: A coed relating a low incident

TABLE 5

COMPARISON OF THE MEAN NUMBER OF STATEMENTS RATED 4 AND 5 BY *Ss* TO THE MEAN NUMBER OF STATEMENTS JUDGED RELEVANT BY THE JUDGES

Incident	Group	$\bar{X}$
High	*Ss*	8.08
	Judges	2.74
Low	*Ss*	5.41
	Judges	1.79
Total	*Ss*	6.77
	Judges	2.25

---

[6] The uncodable category consisted of three high and eight low incidents where either no statements were rated as relevant by the judges or there was no agreement on any statement as being relevant.

describes a time when she started dating other male students and came to the realization that she perhaps didn't love her steady boyfriend. Their relationship became tense and distasteful. Soon she couldn't stand to be with him anymore and the relationship was broken. She then rates the following statements 5 (extremely important):

> I did not gain a feeling of worthwhile accomplishment from my work.
>
> The various facets of my work were not challenging.
>
> I felt that I had not learned anything.
>
> I did not enjoy the kind of work I did.

The following statements are rated 4 (very important):

> I was told that I had done a poor job.
>
> I was dissatisfied because my work gave me very little responsibility.

The judges, on the other hand, rated the following two statements as relevant:

> I got along very poorly with a fellow student.
>
> A close bond between myself and someone in my age group was broken.

The *S* rated both of the above two statements 3 (rather important).

It could be argued that the items rated are, indeed, relevant to the respondent. For example, within the context of the event, the coed certainly did not enjoy the kind of work she did or perhaps she felt that she had done a poor job of breaking up with her boyfriend. The respondent reinterprets items in order to make them seemingly relevant to the incident while the researcher categorizes the same items to work-related behavior because he must stick to the manifest meaning of the items. In this instance, an interpersonal event would lead to a work itself classification.

A concise summary of the findings follows:

1. The traditional M-H theory analysis has resulted in a replication with this student population.

2. The analysis of the ratings by *S*s without considering relevancy did not follow the M-H theory predictions.

3. The analysis of the statements judged relevant resulted in a replication of the M-H theory predictions.

DISCUSSION AND CONCLUSIONS

It is assumed that rating scale procedures will lead *S*s to utilize the lower end of rating scales to avoid the unreasonable position of having to rate the importance of items which are in fact irrelevant to the experiences they are describing. However, the evident fact is that the presence of the low rating scale points does not prevent *S*s from rating as important, items that are actually irrelevant to their experiences. It is difficult to even conceive of any experience to which 5–10 factors could be relevant. The *S*s have to misinterpret or reinterpret the statements in order to produce the exaggerated rating of factors.

When faced with an item, *S*s seem to have a great deal of difficulty in focusing back to the incident and determining what is important in the incident as opposed to what is important to them generally. The force of social desirability or the pull of their value systems is stronger than the directions. Hence, quite automatically, *S*s rate the general importance of the item to them rather than the importance of the item to the incident they have described.

It is not surprising, therefore, that Malinowsky and Barry (1965), Friedlander (1965), and others find the intervening variable of job level affecting the importance of extrinsic and intrinsic factors. What they are measuring with their rating scales are the value systems or the social desirability of blue-collar and white-collar workers.

M-H theory states that both motivator and hygiene factors are equally important for human beings. However, more important, the theory indicates the feelings that result in individuals when certain psychological processes are in operation. It describes the underlying dynamics that are obscured by manifest value systems and social desirability of item wording.

In summary, the rating scale techniques lead to inappropriate and misleading results when used to examine either the M-H theory or job attitudes in general. Unless the problems of methodological coercion forcing irrelevant ratings, overrating by *S*s, and misinterpretation or reinterpretation of items by *S*s can be corrected, the use of these techniques in job attitude research will remain a stumbling block to understanding job attitudes.

REFERENCES

BURKE, R. J. Are Herzberg's motivators and hygienes unidimensional? *Journal of Applied Psychology*, 1966, **50**, 317–321.

DUNNETTE, M. Factor structure of unusually satisfying and unusually dissatisfying job situations for six occupational groups. Paper presented at a meeting of the Midwestern Psychological Association, Chicago, April–May 1965.

DUNNETTE, M., CAMBELL, J. P., & HAKEL, M. D. Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Organizational Behavior and Human Performance*, 1967, **2**, 143–174.

DUNNETTE, M., & KIRCHNER, W. *Psychology applied to industry*. New York: Appleton-Century-Crofts, 1965.

EWEN, R. B., SMITH, P. C., HULIN, C. L., & LOCKE, E. A. An empirical test of the Herzberg two-factor theory. *Journal of Applied Psychology*, 1966, **50**, 544–550.

FRIEDLANDER, F. Job characteristics as satisfiers and dissatisfiers. *Journal of Applied Psychology*, 1964, **48**, 388–392.

FRIEDLANDER, F. Comparative work value systems. *Personnel Psychology*, 1965, **18**, 1–20.

GRAEN, G. B. Motivator and hygiene dimensions for research and development engineers. *Journal of Applied Psychology*, 1966, **50**, 563–566.

HERZBERG, F. *Work and the nature of man*. Cleveland: World, 1966.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. *The motivation to work*. New York: Wiley, 1959.

MALINOVSKY, M. R., & BARRY, J. R. Determinants of work attitudes. *Journal of Applied Psychology*, 1965, **49**, 446–451.

ROSEN, H. Occupational motivation of research and development personnel. *Personnel Administration*, 1963, **26**, 37–43.

STERN, M. L. An investigation of slippage in the motivation-hygiene theory. Unpublished doctoral dissertation, Case Western Reserve University, 1968.

WERNIMONT, P. F. Intrinsic and extrinsic factors in job satisfaction. Unpublished doctoral dissertation, University of Minnesota, 1964.

WERNIMONT, P. F. Intrinsic and extrinsic factors in job satisfaction. *Journal of Applied Psychology*, 1966, **50**, 41–50.

WHITSETT, D. A., & WINSLOW, E. K. An analysis of studies critical of the motivator-hygiene theory. *Personnel Psychology*, 1967, **20**, 391–415.

(Received January 15, 1970)

ournal of Applied Psychology
971, Vol. 55, No. 1, 80-86

NEW APPROACH TO THE PREDICTION OF MEDICAL SPECIALIZATION:

STUDENT-BASED STRONG VOCATIONAL INTEREST BLANK SCALES

GARY T. ATHELSTAN[1] AND GERALD J. PAUL

Center for Research and Education in the Health Occupations, American Rehabilitation Foundation, Minneapolis

In an attempt to devise means of predicting eventual specialization for medical students, longitudinal data from 1,583 *Ss* were studied. Specialty membership was determined in a follow-up 11 yr. after the Strong Vocational Interest Blank (SVIB) was administered. SVIB scales for surgeons, obstetricians, pediatricians, and psychiatrists were developed, using the items which differentiated the specialist groups as freshmen. Senior-based scales were also tested, but predicted no better than those using freshman responses. The resulting student-based scales were reliable on retest, but on measures of distributional overlap, they did not perform as well as standard SVIB occupational scales. However, their predictive validity was superior to that of previously developed SVIB scales for medical specialists.

Some 20 yr. ago, under a contract with the United States Army and the Surgeon General's office, Strong and Tucker (1952) developed the Strong Vocational Interest Blank (SVIB) scales for surgeons, internists, pathologists, and psychiatrists. Their criterion groups consisted of 1,500 established medical specialists who were diplomates in their respective specialties, while 500 physicians-in-general comprised the reference group. The ensuing scales correctly identified, on the average, 81% of the specialists; that is, an individual generally earned his highest score on the scale for his specialty. The average scale intercorrelation was −.17, and the scales held up well on cross-validation. In general, the scales appeared very promising for their intended uses in allocating drafted physicians to different specialty residencies in the armed forces, and in career-choice counseling of medical students.

As part of the study, the SVIBs of 783 medical students were gathered for the purpose of later follow-up to determine how well the scales predict eventual specialty choice. When these students were followed-up in 1960 (Tucker & Strong, 1962), it was found that their medical specialist scores did not correspond with their type of practice. The scales did not discriminate the specialist groups, and the majority of the 1950 students earned their highest score on the psychiatrist scale. Recently, Campbell (1966) completed a reanalysis of all the original data. No significant errors were found.

In attempting to account for the failure of the scales to predict specialization, Tucker and Strong (1962) suggested that changes in the criterion groups themselves could be a major factor. In 1950, when the original criterion groups were studied, only about 15% of physicians were certified specialists. By 1960, when they followed-up their 1950 students, the rate of claimed specialization had risen to about 75%, and 47% of all physicians were actually certified in a specialty. Strong and Tucker hypothesized that medical students began specializing for different reasons than those prevailing prior to 1950. If interests still played a major role in determining specialty choice, the patterns of interest associated with a particular choice may have changed substantially.

Another explanation hinted at by Tucker and Strong (1962), and reinforced by Campbell (1966), suggests that measurable specialty interests may not appear until the physician has been in specialized practice for a time; that is, interests may be modified by membership in a specialty. If this were substantially true, of course, it would never be possible to predict specialization with sufficient accuracy to make the effort practically worthwhile.

---

[1] Requests for reprints should be sent to Gary T. Athelstan, Director of Research, Center for Research and Education in the Health Occupations, American Rehabilitation Foundation, 1800 Chicago Avenue, Minneapolis, Minnesota 55404.

Although it may be that the fully developed interests of a mature specialist do not take their final form until after several years of practice, the characteristics that eventually lead a physician to specialization may already be present and measurable in a student, but in a different, or possibly immature form. Indeed, the manifest differences between such diverse medical specialists as, for example, surgeons and psychiatrists, seem great, not only in interests but in personality variables as well. Such differences should certainly be apparent in some measurable degree in students who are future psychiatrists and surgeons. Furthermore, the differences among students, if measurable, may be of a more basic nature, and therefore less subject to the rapid changes which seem to occur in the adult specialist groups.

The problems to which Strong and Tucker's (1952) work was addressed are even more pressing today. Well over 90% of new physicians now plan to specialize, and the student may have to embark on a somewhat differentiated training program as early as the third year in medical school. While the pressures to make early specialty choices are increasing, there is no corresponding increase in opportunities for students to gain early exposure to the various specialties. This has created a situation which is very stressful for students. "Insecurity regarding professional future" was mentioned as a factor interfering with medical school progress by 42% of a large sample of medical students (Johnson & Hutchins, 1966), the most important of nearly two dozen such factors which were studied.

Although there has been little specific study of career satisfaction among medical specialists, there is anecdotal and indirect evidence that dissatisfaction with specialty choice may also be an important problem. Significant numbers of specialists in some fields (e.g., Athelstan, 1968; Miller, Coker, Greenberg, & McConnell, 1966) either report deep dissatisfaction or demonstrate it by changing specialties.

The present investigation was undertaken in the belief that the problems of predictin and/or assisting the specialty choices of physicians are important enough to warrant ye another attempt to apply the SVIB to tha end. Since the earlier work in this area seemed to have ruled out the usefulness of the traditional SVIB approach to the measurement o medical specialists' interests, this study employed a different method: The specialis scales were based on the SVIB items which differentiated students who later became specialists; that is, the responses of the student were used, rather than those of mature specialists.

METHOD

Subjects

The *S*s were 2,401 freshmen comprising nearly th entire entering classes of a representative, stratifie sample of 24 medical schools (Hutchins, 1964). The were tested with the SVIB and other instruments i 1956 and again in 1960 as part of the longitudina study of medical students conducted by the Associ ation of American Medical Colleges (AAMC) Subsequent specialization was determined by th AAMC as of 1967, in a review of American Medica Association records.

In order to compare directly the effectiveness o freshman-based and senior-based scales, the analyse were restricted to 1,583 *S*s who had both 195 (freshman) and 1960 (senior) SVIBs, and who late confined their specialization, if any, to one fiel The original sample of freshmen was reduced b 617, a figure which included those *S*s who left scho before graduating; who were not seniors in 1960; o who, for various other reasons, were not availab for the retest. An additional 205 *S*s were not i cluded in the analyses because they reported two o more areas of specialization upon follow-up. Th rate of specialization for the final sample was abou 80%.

Procedure

The 1,583 *S*s were grouped according to specialt reported in 1967. The SVIB item response compari sons were made for each specialty group, with a other *S*s serving in turn as the reference group. As standard procedure in the construction of SVI scales, items were selected for inclusion in the scal on the basis of how well they differentiated th responses of the specialist or criterion groups fro

---

[2] The authors gratefully acknowledge the assistan of Edwin B. Hutchins, now professor in the Resear Office of the Vice President for Student Affairs Iowa State University, and David P. Campbell, D rector of the Center for Interest Measurement R search at the University of Minnesota, in obtaini and working with these data.

TABLE 1

NUMBER OF ITEMS SHOWING RESPONSE DIFFERENCES OF TEN PERCENT OR MORE IN ITEM RESPONSE COMPARISONS BETWEEN THE SPECIALTY CRITERION GROUPS AND ALL OTHER *S*s

Group	General practice	Internal medicine	General surgery	Obstetrics-Gynecology	Psychiatry	Pediatrics
Freshman criterion vs. reference group	9	7	24	56	57	44
Senior criterion vs. reference group	12	16	33	64	93	49

Note.—Maximum possible = 288.

the responses of all others. For example, if 70% of future surgeons responded "like" to the school-subject item physiology, while only 50% of other future physicians responded similarly, that item would be said to yield an item response difference of 20% between surgeons and physicians-in-general; it would therefore be an appropriate item to include in the surgeons' scale. To be included in a scale, an item must always differentiate at a statistically significant level, at least. Usually, a minimum of 10% item response difference is required, even if statistical significance is reached below 10%. It is considered desirable to have all of the items in a scale produce response differences of 15% or higher, but items of lower power are acceptable, if they are needed to provide a sufficient number of items for an adequate scale.

In order to develop scales which might be used with the 1966 revision of the SVIB, these analyses were confined to the 288 items common to the past and present forms. Initially, two or more scales were developed for each criterion group, at least one based on the responses of each freshman group, and at least one for each of the groups as seniors. The minimum level of response difference of any of the items used in scales was 10%.

All scores on the scales thus developed were then standardized, giving the criterion group a mean of 50 and a standard deviation of 10. Every SVIB was scored on the freshman specialty scales, the senior specialty scales, and a graduation scale, which was based on items differentiating between all freshmen and all seniors. Group means, scale intercorrelations, test-retest correlations, distributional overlap and predictive hit rates were calculated for each scale. Since we were especially concerned with scale validity, particular emphasis was placed on the hit rates, that is, the proportion of *S*s whose highest score was on the scale corresponding to their specialty.

TABLE 2

DESCRIPTIVE CHARACTERISTICS OF THE STUDENT-BASED SCALES FOR FOUR MEDICAL SPECIALTIES

Student-based scales	Specialty			
	General surgery	Obstetrics Gynecology	Psychiatry	Pediatrics
Freshman				
No. of items per scale	24	45	45	44
Minimum item response difference (%)	10	11	11	10
Senior				
No. of items per scale	33	41	45	49
Minimum item response difference (%)	10	12	14	10

RESULTS

There were six specialty areas with sufficient representation to serve as possible criterion groups. These were: general practice ($n = 277$), general surgery ($n = 166$), internal medicine ($n = 257$), obstetrics-gynecology ($n = 86$), pediatrics ($n = 87$), and psychiatry ($n = 113$). As shown in Table 1, the comparisons of the item responses of general practitioners and of internists with those of all others did not yield a sufficient number of differentiating items to consider developing scales for those specialties. This was true of the responses of those groups as seniors, as well as for the freshmen; consequently, further analyses were confined to the remaining four specialty groups. The number of items differentiating the surgeons (GS), the obstetricians (OB), the psychiatrists (PY), and the pediatricians (PD) from all others, while not great, seemed sufficient for use in experimental scales. Since there was a uniform tendency for the seniors to produce more differentiating items than the freshmen, scales were constructed for both levels.

The scales were designed to have 40–50 items, with the items having the maximum level of item response difference possible at that scale length. Table 2 shows that all scales but GS met the desired criteria of length and level of item response difference. Though it contained only 24 items with response differences of 10% or greater, and thus did not meet the length criterion, the freshman GS scale performed satisfactorily in all subsequent analyses.

These initial scales were evaluated by determining what percentage of each criterion group scored higher on its own scale than on the other three scales. Both the freshman and the senior scales performed adequately on this measure, with 56% of the specialists scoring highest on their own scale. Since the overall performance of the two sets of scales was very similar in this and subsequent analyses, with the freshman scales usually slightly better despite their shorter length, further study of the senior-based scales is not included in this report.

Although these proportions of *S*s scoring highest on their own scales are much lower than those obtained by Strong and Tucker (1952), it must be pointed out that their rate of 81% was based on an analysis of only two scales at a time. A similar evaluation of the student-based scales resulted in these scales correctly identifying 75% of the criterion groups.

The standard adult PY scale was scored for comparison along with the student-based PY scale; the placement rate of the student-based scale was somewhat superior. While 60% of the freshman PY group earned their highest score on their own scale, when the adult PY scale was substituted, the figure declined to 46%. The two PY scales had only 11 items in common, comprising slightly over 20% of the total items in each scale, but they correlated .58 with each other in the entire sample of freshmen.

The graduation scale was devised in the hope that its relationship to other scales or to sample characteristics would reveal something of the nature of interest development or of interest stability among the students. This it did not do. However, the content of the scale, that is, the items to which the seniors responded differently from the freshmen, is worthy of mention. The scale contained 31 items with response differences above 10%; on 28 of these items, the difference was in the direction of decreased liking on the part of the seniors. Nearly all of the items which lost

TABLE 3

MEAN SCORES OF FOUR SPECIALTY GROUPS AS FRESHMEN AND AS SENIORS ON FRESHMAN-BASED SCALES

Specialty groups	Freshman-based specialty scales				Average other specialty
	General surgery	Obstetrics-Gynecology	Pediatrics	Psychiatry	
Freshmen					
General surgeons	50	40	34	41	38
Obstetricians	43	50	33	38	38
Pediatricians	40	38	50	42	40
Psychiatrists	38	35	36	50	36
Internal medicine	40	38	36	44	40
General practitioners	43	38	37	39	39
Other physicians	41	38	35	41	39
Seniors					
General surgeons	48	40	30	40	37
Obstetricians	42	46	30	40	37
Pediatricians	39	36	40	42	39
Psychiatrists	37	33	34	50	35
Internal medicine	39	37	34	45	39
General practitioners	44	39	32	39	39
Other physicians	41	37	32	42	38
4-yr. test-retest correlations	.67	.72	.65	.77	—

TABLE 4

RELATIONSHIP OF PREDICTED SPECIALIZATION TO ACTUAL SPECIALIZATION

Actual specialization	Predicted specialization											
	General surgery			Obstetrics-Gynecology			Pediatrics			Psychiatry		
	Enter	Not enter	Total	Enter	Not enter	Total	Enter	Not enter	Total	Enter	Not enter	Total
Entered	100	66	166	49	37	86	42	45	87	68	45	113
%	7	4	11	3	3	6	3	3	6	5	3	8
Not entered	391	898	1289	289	1080	1369	229	1139	1368	350	992	1342
%	27	62	89	20	74	94	16	78	94	24	68	92
Total	491	964	1455	338	1117	1455	271	1184	1455	418	1037	1455
%	34	66	100	23	77	100	19	81	100	29	71	100

Note.—Total N = 1,455; specialty was not determined for 128 *Ss*.

favor among the seniors had medical or related content, for example, the occupations of chemist; dentist; pharmacist; surgeon; and the school subject, zoology. The greatest decline in like responses was for lab technician, which lost 33% of the *S*s who liked it as freshmen. The only activity which gained proponents among the seniors was bridge-playing; the other two items which changed reflected the increased tendency of the seniors to describe themselves as having mechanical ingenuity and many novel ideas.

Table 3 demonstrates fair separation in the mean scores of the various specialty groups on the scales. The differences on their respective scales between the criterion groups and all others, average slightly over one standard deviation. Each criterion group earns its highest mean score on its own scale, even upon retest as seniors. The PD are an exception, since their mean score as seniors on their own scale is slightly lower than their mean on the PY scale. Table 3 also indicates a moderate level of shrinkage in the scores of the criterion groups on retest (three points, on the average), and a very satisfactory test-retest correlation, which averages .70. The same measure for all of the regular SVIB vocational scales in this group was .67.

Nearly all of the scales intercorrelate negatively; the only exceptions are between GS and OB at .12 and between PD and PY at .11. The mean intercorrelation of all the scales is −.15; this may be compared to the −.17 reported by Strong and Tucker (1952).

The mean overlap between the distributions of the criterion groups and their respective reference groups was 60%, as determined by Tilton's (1937) formula. The poorest scale was GS at 67%, while the PD scale was the best with a percentage overlap of 51%. On the retest, the average overlap increased to 71%. In comparison, the average distributional overlap of the regular SVIB scales is 33%; the corresponding figure for Strong and Tucker's (1952) adult specialist scales was 49%.

In Table 4 the performance of the scales is evaluated in terms of predictive hits and misses for each scale. To the best of our knowledge, this type of analysis has not previously been performed for any SVIB scales. An *S* was predicted to enter the specialty for which he earned his highest score. If he subsequently entered the predicted specialty, a predictive hit was scored and counted as part of the true positive rate. Those *S*s who did not follow the prediction to enter a given specialty were counted as false positives for the scale in question, and entering an unpredicted specialty was called a false negative error.

Table 4 illustrates how the scales predict for the individual specialties. One hundred *S*s

(7%) were predicted to enter GS and actually did so. These *S*s constituted 20% of the 491 *S*s who were predicted to enter that field and 59% of the 166 who actually entered. The true positive hit rate for each scale can thus be obtained by relating the number of correctly predicted entrants into each specialty to the total number of actual entrants. This rate for the OB scale, for example, is 49 out of 86, or 57%; the corresponding figure for PD is 48% (42/87), and for PY it is 60% (68/113).

All scales showed a high false positive rate. This is not especially serious for two reasons. First, all the cell entries are affected in some degree by the number of scales available; in this analysis the scale on which an *S* scores highest is the basis for predicting that he will enter the specialty. Second, for counseling purposes it does no harm to consider several possible specialties; it would be less desirable to incorrectly rule out a specialty which could be an appropriate choice. The false negative rate is also more serious from a technical point of view; while increasing the number of scales in this analysis would probably lower the false positive rate, it could only increase the false negative rate.

Since this analysis does not take into account the absolute level of any of the scores, the figures may overestimate the amount of error which could be anticipated in the counseling use of the scales. Nevertheless, even when used in this oversimplified prediction paradigm, the performance of the scales appeared to provide a significant improvement on the accuracy of chance predictions.

SUMMARY AND DISCUSSION

In an attempt to determine whether the interests associated with specialization in medicine can be identified early enough to assist the career choices of medical students, a new method of constructing SVIB medical specialist scales was tested. Scales for GS, OB, PD, and PY were developed, using SVIB items which differentiated medical students grouped on the basis of membership in those specialties 11 yr. after completing the SVIB. Internists and general practitioners were also studied, but their SVIB item responses did not differ sufficiently from those of other physicians to permit the development of scales.

Although the performance of the resulting scales was considerably poorer than that of the standard SVIB occupational scales, they demonstrated the possibility of predicting specialization from scales based on the responses of medical students. These student-based scales may have greater predictive validity than the medical specialist scales developed by standard SVIB scale construction techniques. On such measures as scale intercorrelations, distributional overlap, and predictive hit rates, the student-based scales performed about as well as the adult specialist scales previously developed by Strong and Tucker (1952). The 4-yr. test-retest reliability of the student-based scales was nearly identical to that of standard SVIB scales scored for the same sample.

The use of the student-based scales in counseling medical students is not clearly indicated at this time, although they certainly permit greater accuracy of prediction than that obtainable by chance, which seems to be the primary alternative means of prediction. Further research on these scales is needed for the following reasons: In the first place, in order to be of much practical value the coverage of specialties should be increased. It should prove possible, for example, to develop similar scales for such major fields as anesthesiology, ophthalmology, pathology, and radiology. Secondly, the validity of these scales should be further tested. We hope to do this in a very direct way in the near future, by evaluating the predictive validity of the scales for other samples of physicians on which longitudinal data can be obtained.

This investigation has demonstrated the feasibility of predicting specialization from medical school freshman SVIB responses. However, it is not likely that any amount of refining of techniques would result in significant improvement on the level of accuracy obtained here. The primary reasons for this are the wide range of job functions found in each specialty, and the fact that interests probably covary more with the functional dimensions in medicine than with the formal specialty divisions. The interests of physicians engaged in such activities as research and teaching, administration, or straight clinical practice probably resemble the interests of their colleagues in those functional areas, irrespective of specialty membership, more closely than they resemble the interests of their respective cospecialists. Several studies (e.g., Schumacher, 1964) support this hypothesis, and the future research program of the Center for Research and Education in the Health Occupations will deal specifically with that question in a follow-up of the *S*s studied here.

REFERENCES

ATHELSTAN, G. T. *The vocational interests, values, and career development of specialists in physical medicine and rehabilitation.* (Bulletin No. 9 of the Commission on Education in Physical Medicine and Rehabilitation) Minneapolis, 1968.

CAMPBELL, D. P. *Re-analysis of Strong's interest data from medical specialists.* (Final Report for U. S. Office of Education Contract No. OE-6-10-148, Project No. 5-8404) December 1966.

HUTCHINS, E. B. The AAMC longitudinal study: Implications for medical education. *Journal of Medical Education,* 1964, **39**(3), 265–277.

JOHNSON, D. G., & HUTCHINS, E. B. Doctor or dropout? *Journal of Medical Education,* 1966, **41**, 1099–1272.

MILLER, N., COKER, R. E., JR., GREENBERG, B. G., & MCCONNELL, F. S. Toward a typology of public health careers. *Milbank Memorial Fund Quarterly,* 1966, **46**, 200–213. (Pt. 1)

SCHUMACHER, C. F. Personal characteristics of students choosing different types of medical careers. *Journal of Medical Education,* 1964, **39**, 278–288.

STRONG, E. K., JR., & TUCKER, A. C. The use of vocational interest scales in planning a medical career. *Psychological Monographs,* 1952, **66** (9, Whole No. 341).

TILTON, J. W. The measurement of overlapping. *Journal of Educational Psychology,* 1937, **28**, 656–662.

TUCKER, A. C., & STRONG, E. K., JR. Ten-year follow-up of vocational interest scores of 1950 medical college seniors. *Journal of Applied Psychology,* 1962, **46**, 81–86.

(Received January 15, 1970)

Manuscripts Accepted for Publication in the *Journal of Applied Psychology*

Screening Employment Applicants for Attitudes Toward Theft: Philip Ash*: Department of Psychology, University of Illinois, Box 4348, Chicago, Illinois 60680.

Note on the Methodology of the Graen, Orris, and Alvares Studies Testing the Contingency Model: Fred E. Fiedler*: Organizational Research, University of Washington, Seattle, Washington 98105.

Contingency Model of Leadership Effectiveness: Some Methodological Issues: George Graen*, James Burdeane Orris, and Kenneth M. Alvares: Institute of Labor and Industrial Relations, University of Illinois, Champaign, Illinois 61820.

Method for Detecting Random Answers on Personality Questionnaires: Jerry W. O'Dell*: Associate Professor of Psychology, Eastern Michigan University, Ypsilanti, Michigan 48197.

Effectiveness of Two Orientation Approaches in Hard-Core Unemployed Turnover and Absenteeism: Hjalmar Rosen*, and John Turner: Department of Psychology, Wayne State University, Detroit, Michigan 48202.

General Aptitude Test Battery Scores for Men Differing in Age and Socioeconomic Status: James L. Fozard*, and Ronald L. Nuttall: Research Psychologist, Veterans' Administration Outpatient Clinic, 17 Court Street, Boston, Massachusetts 02108.

Effect of Job Attitudes, Training, and Organization Climate upon Performance of the Hard-Core Unemployed: Frank Friedlander*, and Stuart Greenberg: Department of Management Studies, University of Leeds, Leeds LS2 9JT, England.

Morale and Productivity Consequences of Group Leadership Style, Stress, and Type of Task: Leonard L. Rosenbaum*, and William B. Rosenbaum: Department of Psychology, Montgomery College, Rockville, Maryland 20850.

Effects of Target Value and Exposure Duration upon Recall in a Visual Search Task: Harvey Taub*, and Charles Abrams: Chief, Research Section, Psychology Service, Veterans' Administration Hospital, Irving Avenue and University Place, Syracuse, New York 13210.

Motivation in Managerial Levels: The Relationship of Need Satisfaction to Job Performance: John W. Slocum, Jr.*: Associate Professor of Management, Pennsylvania State University, 223 Boucke Building, University Park, Pennsylvania 16802.

---

* Asterisk indicates author for whom address is supplied.

(Continued on page 94)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 87–91

EFFECT OF PRICE INFORMATION ON PRODUCT QUALITY RATINGS

E. R. VALENZI AND I. R. ANDREWS[1]

Bowling Green State University

Three food products discriminable in a blind-taste test condition and varying widely in price were rated on taste quality by 46 *S*s. Each of the three products was identified twice as a low-price product, twice as a medium-price product, and twice as a high-price product. Confirming two hypotheses, taste quality ratings were positively related to price information given and to actual product differences. Contrary to the third hypothesis, actual product differences accounted for approximately three times as much variance as price information given. Several differences between the laboratory setting and actual consumer behavior may have led to an underestimation of the effects of price cues on consumer quality judgments.

Several studies have evaluated the influence of information on consumers' quality judgments. Allison and Uhl (1964) showed that different brands of beer undistinguishable in the blind taste-test condition were rated significantly different when brand names were provided. Makens (1965) presented *S*s with samples of turkey cut from the same bird but identified as a known brand name and an unknown brand name. The sample labeled with the known brand name was rated superior on a hedonic scale. Petit (1958) showed that information could affect consumer preference ratings if it was meaningful to the rater. Raffensperger and Pilgrim (1956) found that information affected consumers' discriminations even when it contradicted perceptual experience. These studies indicate that, besides perceptual experience with the product itself, product information has a significant influence on consumer quality judgments.

Price is frequently suggested as an important variable affecting perception of product quality. Leavitt (1954) and Tull, Boring, and Gonsior (1964) used price as an independent variable with relatively homogeneous products and found a positive price-quality relationship. Their experimental task was a paper-and-pencil one and did not allow *S*s to use the products evaluated. McConnell (1968) presented the same brand of beer in bottles labeled with one of three relatively meaningless symbols (L, M, P) to represent brand names. Each brand was labeled with a different price and consumers chose a brand, drank the beer, and then rated it on several characteristics. Ratings were positively related to the price information provided, demonstrating that price was a cue influencing the perception of quality.

All the studies cited above used products that were homogeneous in at least two ways: (*a*) the products were identical or highly similar so that they could not be discriminated in a blind condition, and (*b*) the products did not vary widely in purchase price. The results of these studies are therefore limited in generality to nondiscriminable products that are used for the same purpose and are approximately equivalent in price.

White (1966) reported an experiment that used products discriminable in a blind condition. Two brands of adding machines were evaluated by users under three conditions: blind, correct brand labeling, and with brand labels reversed. In the blind condition the unknown brand was rated somewhat superior, in the correct labeling condition the unknown brand was rated as somewhat inferior, and when the labels were reversed the machine with the known brand label was rated far superior. These results indicate that even when there is a difference between products, quality ratings can be reversed by product information.

The present experiment extended previous research by testing three products used for the same purpose that varied widely in price and were discriminable in the blind-test condition. In addition, an experimental design was employed that compared the effects of price information against the effects of actual product differences (i.e., differences in quality ratings measured in a blind-test condition).

---

[1] Requests for reprints should be sent to I. Robert Andrews who is now at the Department of Economics and Commerce, Simon Fraser University, Burnaby 2, British Columbia, Canada.

It was hypothesized that quality ratings would be positively related both to price information given and to actual product differences, but that price would have the larger effect. It was hypothesized that price would have a larger effect because actual product differences identified in a previous experiment (Andrews & Valenzi, 1970) were relatively small while the price information given included substantially different prices.

METHOD

Subjects

Seventy-seven female, introductory psychology volunteers were divided into two groups, an Information group ($n = 46$) that received correct and incorrect price information, and a Blind group that received no price information ($n = 31$) other than the range in purchase price of the products used. The No Information group was included only to verify that products were discriminable in the blind-test condition.

For the Information group, one of the two independent variables was actual product differences. This was defined by choosing three products, Ever-Sweet butter at 95¢ a pound, Imperial margarine at 45¢ a pound, and Nutley margarine at 20¢ a pound. EverSweet butter and Nutley margarine are local brands, while Imperial margarine is a national brand. The blind-test quality ratings of these products were significantly different when measured with a paired comparison design (Andrews & Valenzi, 1970). As a preliminary part of the present study these blind-test quality rating differences were confirmed with a single stimulus method of presentation. For the 95¢, 45¢, and 20¢ a pound spreads, respectively, the blind-rated quality means (on a 9-point scale) were 5.82, 5.39, and 4.66. Though highly significant, $F = 10.7$, $df = 2/60$, $p < .01$, these differences in mean quality ratings were surprisingly small when compared to the substantial differences in price per pound. A post hoc analysis using the Newman-Keuls procedure (Winer, 1962, pp. 80–85) showed that the means for both the 95¢ spread and the 45¢ spread were significantly different from the 20¢ spread mean at the .01 level. The difference between the 95¢ and 45¢ spread means was not significant in this sample. In an earlier study using the paired comparison method of stimulus presentation, this difference was significant at the .05 level.

The second independent variable was price information given. The *S*s tasted each of the three products six times, twice identified as a high-priced spread, twice identified as a medium-priced spread, and twice identified as a low-priced spread. For both groups, order of stimulus presentation was determined randomly for each *S*.

Procedure

The products were presented to *S*s on 1-in. squares of plain white bread, spread somewhat more thickly than usual to facilitate tasting. In addition, each square of bread was placed spread-side down on a tray lined with aluminum foil. The *S* was instructed to place the bread spread-side down on his tongue without looking at the spread. This procedure minimized visual cues, especially color (the Nutley margarine was noticeably lighter), and facilitated tasting.

Each *S* received an aluminum tray with 18 squares of bread in three rows. For the Information group, the price per pound of the spread was shown on a strip of tape placed immediately above the bread. The *S*s were told that they would be tasting 18 different spreads purchased from six different stores in the area. For this reason 18 different prices were used, 6 high (95¢, 96¢, 98¢, \$1.01, \$1.03, \$1.04), 6 medium (36¢, 38¢, 39¢, 42¢, 43¢, 45¢), and 6 low (13¢, 14¢, 16¢, 18¢, 21¢, 22¢). These prices were chosen because they were realistic and represented the actual range of prices for these products in local stores and markets.

As part of the introduction and instructions to the Information group, *S*s were told that this was the second part of a two-part experiment, that the first part used a sample of housewives as judges, and that our purpose in this part of the experiment was to see if results obtained from college students were comparable to those of the general population. To decrease the chance that *S*s would deduce the price-quality hypothesis of the experiment, they were told also that the overall purpose of the study was "to see if rinsing the mouth between tastes was a critical variable in sensory taste-testing experiments." To lend credibility to the deception, the instructions required *S*s to take a sip of water between tastes on half of the trials and swish it around in their mouth for 5–10 sec. before swallowing. Half of the *S*s rinsed on the first 9 trials and half on the second 9 trials to further establish the notion that the rinse was the independent variable. To explain the presence of the price information immediately above the stimuli, it was stated that many of the housewives in the first part of the study wanted to know the cost or brand names of the products they were tasting and that this information was provided to ensure continued cooperation and high motivation. To keep conditions comparable for the two samples, the price information had to be provided for the college student sample. Postexperimental questionnaire data bearing on the success of the instructions is presented in the Discussion section.

The interval between tastes was set at 2 min. and timed with a stopwatch. The relatively long intertrial interval was chosen to allow the natural salivary processes to remove the residual food particles before the next taste.

After tasting each sample, *S*s rated it on a 9-point taste quality scale identical to Bruvold's (1968) vertical, equal interval scale for rating the taste of water except that the word, water, was changed to the word, spread. The scale ranged from 1, "this spread has a horrible taste" to 9, "this spread has an excellent taste."

RESULTS

Each of the nine different combinations of three actual products paired with three different levels of price information was rated twice by each of the 46 *S*s in the Information group. The average of these two ratings for each of the nine product/price combinations became the basic data for each *S*. These basic data were analyzed in a 3 × 3 × 46 analysis of variance design with one *S* per cell (Lindquist, 1956, pp. 237–238).

As shown in Table 1, the main effects for both independent variables were significant. Inspection of the cell means in Table 2 showed that, with only one exception (the inversion of means 4.59 and 4.67), differences between adjacent cells were in the expected direction. As predicted, quality ratings were positively related to price information given and also positively related to actual product differences despite the presence of the price information.

Not conforming fully to the expected pattern of results was the low-priced margarine which received a higher rating when identified as a low-priced margarine than when it was identified as a medium-priced margarine. As explained below, this slight and nonsignificant reversal might have been an artifact of the rating scale. Also a possible artifact of the rating scale was the significant interaction between the two independent variables (see Table 1).

TABLE 1

ANALYSIS OF VARIANCE SUMMARY TABLE FOR PRODUCT QUALITY RATINGS

Factor	*SS*	*df*	*MS*	*F*	ω^{2}[a]
Actual product (A)	131	2	65.5	22*	.135
Price information (B)	45	2	22.5	10.1*	.041
*S*s (C)	134	45	2.97		
A × B	17	4	4.25	12.9*	.020
A × C	247	90	2.74		
B × C	193	90	2.14		
A × B × C	56	180	.31		
Total	923				

[a] Omega-squared represents an estimate of the proportion of variance accounted for by the independent variable and is given by $\omega^2 = \frac{\sigma^2_y - \sigma^2_y/x}{\sigma^2_y}$.

* $p < .05$.

TABLE 2

MEAN PRODUCT QUALITY RATINGS FOR INFORMATION GROUP

Actual purchase price	Price information given			$\bar{X}$
	95¢–$1.04	36¢–45¢	13¢–22¢	
95¢	6.66	6.07	5.66	6.13
45¢	6.28	5.92	5.51	5.90
20¢	5.26	4.59	4.67	4.84
$\bar{X}$	6.07	5.53	5.28	

The most surprising result was the nonconfirmation of the prediction concerning the relative effects of actual product differences and price information given. Contrary to expectation, the omega-squared values (Hays, 1963, p. 325) in Table 1 showed that actual product differences accounted for 13% of quality rating variance while price information given accounted for only 4%.

Since actual differences in product taste were relatively small and differences in price information given were relatively large, one might be tempted to conclude that the effects of price cues on consumer quality judgments are relatively minor. As discussed below, however, there were several characteristics of this laboratory experiment which might have led to an underestimation of the influence of price.

DISCUSSION

In evaluating the results of this kind of study, it is important to know how *S*s interpreted the purpose of the experiment. Data from a postexperimental questionnaire shed some light on that question. When asked, "What do you think is the purpose of this experiment?", 21 of the 46 Information group

*S*s responded to the effect that they thought the experiment had something to do with price and quality. Eight *S*s guessed an incorrect hypothesis and 17 *S*s accepted the hypothesis implied by the instructions. Of the 21 *S*s who guessed correctly, only one said that she tended to rate according to price rather than on the basis of taste. In addition, the mean product quality ratings for *S*s who guessed correctly were essentially the same as for those who did not. It seems unlikely, therefore, that deducing the correct hypothesis had a profound effect on the quality ratings.

Previous research (Leavitt, 1954; McConnell, 1968; Tull, Boring, & Gonsior, 1964) had shown that product purchase price influences judgments of product quality for nondiscriminable products of approximately equivalent price. The present results may extend that finding to relatively inexpensive food products which are discriminable and which vary somewhat in price.

It is clear, on the other hand, that the price cues given to *S*s did not completely obscure the rather subtle differences in product tastes. In fact, when averaged out across the high-, medium-, and low-price cues given for each product, the row means shown in Table 2 are very similar to the mean quality ratings found when products were tasted under the blind condition.

Since the actual taste differences among products were fairly small, while the differences in price information given were quite large, it was surprising to find actual product differences accounting for three times as much of the dependent variable variance as the price information given. There are several factors which might account for this unexpected result. First, even though price information differences were quite large percentagewise, in an absolute sense there is not too much difference between 20¢, 45¢, and 95¢. Perhaps a stronger price information effect would be obtained with more expensive products, for example, three diamonds whose purchase prices were identified as $200, $400, and $800.

Second, a sample of college girls might have responded without considering the long range impact of the price differences as a housewife would be more likely to do. In other words, a housewife in deciding between products priced at a 75¢ differential might calculate what this price difference adds up to over the course of a year on a frequently purchased product. A college girl, on the other hand, might treat the decision as a one time purchase and therefore view the price difference as trivial.

Also of possible importance is the fact that the present *S*s did not have to make a purchase decision and pay out their own money to back up that decision. Therefore, there was no need to distort product quality evaluations to justify a purchase decision. When a lady spends $40 for an ounce of "royal queen bee jelly," her perception of its cosmetic effect is apt to be distorted.

Also relevant is the fact that these product tastes were compared to one another with only a short time interval between tastes. The typical consumer is more apt to base his product quality comparison on taste samples widely separated in time, thus reducing the accuracy of such comparisons.

Last, it is possible that the effect of price information given was reduced by what might be called a basement effect. As shown in Table 2, there was little difference and even a nonsignificant reversal between the two lowest cell means—4.59 when the low-priced product was identified as a medium-priced product versus 4.67 when the same product was identified as a low-priced product. Apparently, many *S*s were unwilling to say that a neutral tasting product has a "slightly bad taste" even when it was identified as a low-priced product. A different quality rating scale, perhaps a 7-point scale going from neutral to excellent, would eliminate the basement effect and possibly lead to a somewhat greater price information effect. The authors believe, however, that such a scale revision would not alter the results to the point where the relative effect prediction would be supported with the products used in this study.

To summarize, it seems likely that there were several features of our laboratory situation which led to an underestimation of the effect of price on consumer quality judgments. At least it seems very safe to conclude that the effects of price were not overestimated by the results of this study.

To extend the above results, the experiment should be replicated with products other than foods or beverages and with products which have larger absolute price differences. Laboratory studies must be cross-validated by field studies. Critical to a true understanding of the effect of price on product quality evaluations would be an experimental design which allowed consumers to spend their own money in choosing among various priced products. Post-purchase product evaluations would have to be sampled shortly after purchase and again after a longer time interval. Also of interest would be studies which compared price information against brand information to see which has the stronger effect on product quality judgments.

REFERENCES

ALLISON, R., & UHL, P. Effect of brand identification on perception of product quality. *Journal of Marketing Research,* 1964, **1**, 36–39.

ANDREWS, I. R., & VALENZI, E. R. The relationship between product purchase price and blind rated quality: Margarines and butters. *Journal of Marketing Research,* 1970, **7**, 393–395.

BRUVOLD, W. H. Scales for rating the taste of water. *Journal of Applied Psychology,* 1968, **52**, 245–253.

HAYS, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

LEAVITT, H. J. A note on some experimental findings about the meaning of price. *Journal of Business,* 1954, **27**, 205–210.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

MAKENS, J. C. Effect of brand preference upon consumers' perceived taste of turkey meat. *Journal of Applied Psychology,* 1965, **49**, 261–263.

MCCONNELL, J. D. Effects of pricing on perception of product quality. *Journal of Applied Psychology,* 1968, **52**, 331–334.

PETIT, L. A. Informational bias in flavor preference testing. *Food Technology,* 1958, **12**, 12–14.

RAFFENSPERGER, E. L., & PILGRIM, F. J. Knowledge of the stimulus variable as an aid in discrimination tests. *Food Technology,* 1956, **10**, 254–257.

TULL, D. S., BORING, R. A., & GONSIOR, M. A. A note on the relationship of price and imputed quality. *Journal of Business,* 1964, **37**, 186–191.

WHITE, I. S. The perception of value in products. In J. Newman (Ed.), *On knowing the consumer.* New York: Wiley, 1966.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received January 22, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 1, 92–94

RELATIONSHIP BETWEEN JOB ATTITUDES AND TWO FORMS OF WITHDRAWAL FROM THE WORK SITUATION

L. K. WATERS AND DARRELL ROACH

Ohio University

Nationwide Insurance

Attitudes concerning the job in general and specific aspects of the work situation were correlated with termination decisions and recorded frequency of absences during the year following administration of the attitude questionnaire. Overall satisfaction and satisfaction with several intrinsic aspects of the work situation were significantly related to the termination criterion. None of the extrinsic variables were related. For the frequency of absence criterion, attitudes toward specific aspects of the work situation showed no consistent relationships when nonattitudinal variables were partialed out, but overall satisfaction was significantly related.

Vroom (1964) has proposed that

> workers who are highly attracted to their jobs should be subject to stronger forces to remain in them than those who are less attracted to their jobs. These stronger forces to remain should be reflected in a lower probability of behaviors which take the person out of his job, both permanently and temporarily [p. 187].

Reviews of the literature by Brayfield and Crockett (1955); Herzberg, Mausner, Peterson, and Capwell (1957); and Vroom (1964) have generally indicated some relationship between job-related attitudes and both termination and absenteeism criteria. More recent studies by Hulin (1966) and Mikes and Hulin (1968) have supported the hypothesized relationship between attitudes and termination. However, there have been sufficient inconsistencies in the reported relationships that further data, either as essentially replications of previous work or as generalization to different samples and situations, seem warranted. The purpose of the present study was to obtain additional information on the relationships between job related attitudes and (*a*) subsequent termination, and (*b*) recorded frequency of absences.

METHOD

The respondents in this study were 160 nonsupervisory female employees in one regional office of a national insurance company who had responded to a job satisfaction questionnaire as part of an earlier study (Waters & Waters, 1969). The employees ranged in age from the late teens to early 60s and all were at least high school graduates. Immediate supervisors of all employees utilized in the study were also females.

A job attitude questionnaire was administered to small groups of employees by one of the authors during a single working day. Respondents were assured that their individual responses would not be made known to the company. Names were requested but employees were given the option of not responding to that item if they "felt uncomfortable" doing so. Approximately 17% did not give their names.

The job attitude scales were presented in booklet forms and consisted of separate overall satisfaction and dissatisfaction scales (always the first two scales, order randomized), an overall satisfaction/dissatisfaction scale, the five scales of the Job Description Index (JDI; Smith, 1967) and a list of 11 job factors (arranged in alphabetical order) to be rated on a satisfaction/dissatisfaction scale. Ratings of satisfaction/dissatisfaction (both overall and for specific job factors) were made on a 12-point anchored scale and the separate satisfaction and dissatisfaction ratings on 7-point scales which consisted of the appropriate 6 points of the 12-point satisfaction/dissatisfaction scale plus a seventh alternative (not satisfied or not dissatisfied).

Approximately 1 yr. later, termination and frequency of absence data were obtained from company records. Except for 12 employees who had terminated because of pregnancy, all employees leaving the company during the year subsequent to the administration of the job attitude booklets were classified as terminations. No attempt was made to distinguish among these cases with respect to differences in reported reasons for terminating their employment. Absences were recorded as frequency of absences for the year following the attitude booklet administration. For those employees who terminated during the year, frequency of absences while still with the company were adjusted to a per-year frequency. Absence data were not available or could not be matched to job satisfaction booklets in 17 cases. The cases were excluded from the analysis. The final sample consisted of 131 cases, 109 current employees, and 22 terminations.

RESULTS AND DISCUSSION

Table 1 gives the correlations between nonattitudinal variables, overall- and specific job-related attitudes, and the two criterion measures. None of the nonattitudinal variables were related to the termination criterion. Of the overall job attitude measures, those requiring self-reports along a bipolar satisfaction/dissatisfaction or satisfaction only dimension were more related to actual job behavior than reported degree of dissatisfaction. All of the ratings of satisfaction with specific aspects of the work situation that correlated significantly ($p < .05$) with termination were intrinsic variables. These data are consistent with earlier studies of correlates of termination by female clerical workers by Hulin (1966) and Mikes and Hulin (1968) for overall satisfaction measures and by Ross and Zander (1957) for the greater potency of intrinsic variables.

For the frequency of absence criterion, correlations with both job grade and company tenure were significant ($p < .05$). When these variables were partialed out of the attitude-frequency of absence relationships, two of the overall satisfaction measures (satisfaction/dissatisfaction and satisfaction) and two of the specific job factors (JDI work and JDI co-workers) were significantly ($p < .05$) related to frequency of absences. In the literature reviews, only one study (Metzner & Mann, 1953) seemed a reasonably adequate investigation of the relationships between attitudes and absenteeism among female white-collar workers. The data from the present study conflict with the results of the Metzner and Mann study in that they found no relationship between a number of attitudinal items and absenteeism.

The correlation between termination and frequency of absences was −.32 (indicating that those who terminated had more absences than nonterminators). These two forms of withdrawal from the work situation (permanent and temporary) were related to each other and both were related to overall measures of job satisfaction. These results, obtained under conditions in which situational factors appear to have minimal saliency, support Vroom's (1964) proposal that withdrawal from the work situation is related to job attitudes.

TABLE 1

CORRELATIONS BETWEEN PREDICTOR VARIABLES AND TERMINATION AND FREQUENCY OF ABSENCE CRITERIA

Variables	Criteria	
	Termination	Frequency of absences
Nonattitudinal		
Marital status	−.15	.09
Education	−.08	−.02
Job grade	.15	−.23**
Company tenure	.10	−.21*
Job description index		
Work	.24**	−.28**
Pay	.07	−.09
Promotion	.09	.05
Supervision	.14	.05
Co-workers	.11	−.18*
Job factors		
Growth and advancement	.24**	−.07
Recognition for work done	.11	−.10
Responsibility on the job	.38**	−.16
Sense of achievement	.24**	−.18*
Work	.22*	−.20*
Competent supervision	.09	−.09
Considerate supervision	.01	.04
Company policies & practices	.04	−.04
Co-workers	−.06	−.03
Physical working conditions	.01	−.03
Salary	.14	−.21*
Overall measures		
Satisfaction/dissatisfaction	.24**	−.23**
Satisfaction	.23**	−.28**
Dissatisfaction	−.11	.16

* $p \leq .05$.
** $p \leq .01$.

REFERENCES

BRAYFIELD, A. H., & CROCKETT, W. H. Employee attitudes and employee performance. *Psychological Bulletin*, 1955, **52**, 396–424.

HERZBERG, F., MAUSNER, B., PETERSON, R. O., & CAPWELL, D. F. *Job attitudes: Review of research and opinion.* Pittsburgh: Psychological Service of Pittsburgh, 1957.

HULIN, C. L. Job satisfaction and turnover in a female clerical population. *Journal of Applied Psychology*, 1966, **50**, 280–285.

METZNER, H., & MANN, F. Employee attitudes and absences. *Personnel Psychology*, 1953, **6**, 467–485.

MIKES, P. S., & HULIN, C. L. Use of importance as a weighting component of job satisfaction. *Journal of Applied Psychology*, 1968, **52**, 394–398.

ROSS, I. C., & ZANDER, A. F. Need satisfaction and employee turnover. *Personnel Psychology*, 1957, **10**, 327–338.

SMITH, P. C. The development of a method of measuring job satisfaction: The Cornell studies. In E. A. Fleishman (Ed.), *Studies in personnel and industrial psychology*. (Rev. ed.) Homewood, Ill.: Dorsey Press, 1967.

VROOM, V. *Work and motivation*. New York: Wiley, 1964.

WATERS, L., & WATERS, C. Correlates of job satisfaction and job dissatisfaction among female clerical workers. *Journal of Applied Psychology*, 1969, **53**, 388–391.

(Received January 28, 1970)

(Continued from page 86)

Influence of Race on the Manifestation of Leadership: James R. Fenelon, and Edwin I. Megargee*: Department of Psychology, Florida State University, Tallahassee, Florida 32306.

Study of Entrepreneurial Motivation: Comparing the Fast- and Slow-Progressing Small-Scale Industrial Entrepreneurs: P. S. Hundal*: Department of Psychology, Punjab University, Chandigarh, India.

Recognition of Human Faces: Effects of Target Exposure Time, Target Position, Pose Position, and Type of Photograph: Kenneth R. Laughery*, Judith F. Alexander, and Alan B. Lane: State University of New York at Buffalo, Department of Psychology, Buffalo, New York 14214.

Favorable Self-Perception, Perceived Supervisory Style, and Job Satisfaction: Duane E. Thompson*: Assistant Professor, College of Business Administration; University of Iowa, Phillips Hall, Iowa City, Iowa 52240.

Development of the Survey of Work Values: Stephen Wollack, James G. Goodale, Jan P. Wijting, and Patricia Cain Smith*: Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43402.

Motivational Type as a Moderator of the Satisfaction/Performance Relationship: Frank J. Landy*: Department of Psychology, Pennsylvania State University, 417 Psychology Building, University Park, Pennsylvania 16802.

Characteristics of Authors and Funding of Journal Articles in Applied Psychology: Gary N. Chaison*: Department of Industrial Relations, State University of New York at Buffalo, Crosby 319 A, Buffalo, New York 14214.

Measurement of the Distribution of Potential Influence and Participation in Groups and Organizations: Gordon E. O'Brien*, Anthony Biglan, and Judith Penna: Senior Lecturer in Psychology, School of Social Sciences, Flinders University of South Australia, Bedford Park, South Australia 5042.

Task Enjoyment and Task Perseverance in Relation to Task Success and Self-Esteem: Skipton Leonard and Joseph Weitz*: Department of Psychology, New York University, 21 Washington Place, Room 311, New York, New York 10003.

Similarity of Post-Interview Trait Rating Intercorrelations as a Contributor to Interrater Agreement in a Structured Employment Interview: Milton D. Hakel*: Department of Psychology, Ohio State University, 1945 North High Street, Columbus, Ohio 43210.

Leadership Style, Hierarchical Influence, and the Satisfaction of Subordinate Role Expectations: A Test of Likert's Influence Proposition: Robert J. House*, Alan C. Filley, and Damodar N. Gujarati: Bernard M. Baruch College, Baruch Graduate Center, City University of New York, 257 Park Avenue South, New York, New York 10010.

Leadership, Participation, and Self-Orientation in Observed and Non-Observed Groups: Stephen R. Wilson*: Department of Sociology, Temple University, Philadelphia, Pennsylvania 19122.

Identifying Patterns of Teacher Behavior from Student Evaluations: T. W. Field, W. S. Simpkins*, R. K. Browne, and P. Rich: Senior Lecturer in Education; University of New England, Armidale, N.W.S., Australia 2351.

Taylor-Russell Tables for Dichotomous Criterion Variables: Norman M. Abrahams*, Edward F. Alf, and John H. Wolfe: Personnel Measurement Research Department, Naval Personnel and Training Research Laboratory, San Diego, California 92152.

Error and Variability in the Visual Processing of Dental Radiographs: Irwin L. Goldstein*, and William H. Mobley: Department of Psychology, University of Maryland, College Park, Maryland 20742.

Effect of Fittingness, Type of Goods, and Type of Slogan on Brand Awareness: Sasi Misra*, and Subhash Jain: Indian Institute of Management, Vastrapur, Ahmedabad 15, India.

Relative Influence of Job and Supervision on Shared Worker Attitudes: Douglas B. Campbell*: Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

Price, Brand Name, and Product Composition Characteristics as Determinants of Perceived Quality: Jacob Jacoby*, Jerry C. Olson, and Rafael A. Haddock: Department of Psychology, Purdue University, Lafayette, Indiana 47907.

Characteristics of Nonrespondents among Workers: Martin J. Gannon*, Joseph C. Nothern, and Stephen J. Carroll, Jr.: Behavioral Science Division, Department of Business Administration, University of Maryland, College Park, Maryland 20742.

---

* Asterisk indicates author for whom address is supplied.

JOURNAL OF APPLIED PSYCHOLOGY

APRIL 1971 VOL. 55, NO. 2

Contents

ARTICLES

The Culture Assimilator: An Approach to Cross-Cultural Training: *Fred E. Fiedler, Terrence Mitchell, and Harry C. Triandis* 95

Utility of Merit Increases: *Brian A. Giles and Gerald V. Barrett* 103

Effects of Irrelevant Stimuli on the Processing of Information in Complex Displays: *Irwin L. Goldstein, and John C. Allen* 110

Interaction of Environment and Personality: A Test of Holland's Theory: *Charles F. Elton* 114

Job Offer Acceptance Behavior and Reinforcement: *John M. Ivancevich and James H. Donnelly* 119

Display Characteristics in Inspection Tasks: *Robert C. Williges and Harrison Streeter* 123

Pupil Dilation and Its Relationship to Information Processing during Reading and Listening: *Ronald P. Carver* 126

Vocational Interests of Political Scientists: *Jack E. Rossmann, Orville J. Lips, and David P. Campbell* 135

Effects of Temporal Placement of Unfavorable Information on Decision Making during the Selection Interview: *Roger N. Blakeney and John F. MacNaughton* 138

Stability of Vocational Interests within Nonprofessional Occupations: *Alan W. Lau and Norman M. Abrahams* 143

Assimilation of New Members in a Large and a Small Church: *Allan W. Wicker and Anne Mehler* 151

Double Cross-Validation of a Weighted Application Blank over Time: *Darrell E. Roach* 157

Screening Employment Applicants for Attitudes toward Theft: *Philip Ash* 161

Selection and Evaluation Procedure for Operating Room Personnel: *Ury Gluskinos and Thomas F. Brennan* 165

Comparison of Work Groups with Stable and Unstable Membership: *B. D. Fine* 170

LIST OF MANUSCRIPTS ACCEPTED 109, 150

© 1971 by the American Psychological Association

INFORMATION FOR CONTRIBUTORS

Style. A professional article should possess certain characteristics: (*a*) conciseness and an apparent respect for reader time; (*b*) unambiguous and simple vocabulary with technical and erudite words used only when simpler ones would obviously be inadequate; (*c*) conformity to accepted technical style in tables, terminology, and references; (*d*) conclusions that are clearly related to the evidence presented. The reader should be led, step by step, from a statement of problem or purpose, through analysis of evidence, to conclusions and implications. Authors are encouraged to consult the little book entitled *The Elements of Style* by W. Strunk, Jr. and E. B. White (New York, Macmillan, 1962).

Criteria for evaluation. Manuscripts will be evaluated on the basis of several criteria, including: (*a*) significance in contributing new knowledge to the field, (*b*) technical adequacy, (*c*) appropriateness for the *Journal of Applied Psychology*, and (*d*) clarity of presentation.

Format. Manuscripts must be prepared in the format described in the *Publication Manual of the American Psychological Association* (1967 Revision), obtainable for $1.50 from the American Psychological Association, 1200 17th Street, N.W., Washington, D. C. 20036. Articles not prepared in this manner cannot be reviewed. Special attention should be given to the section on typing the manuscript (p. 48) and to the sections on tables, figures, and references. Note that *all copy must be double spaced*, including references, title, figure captions, etc. The senior author's last name should appear on each page in the upper left-hand corner. In reference lists, give journal titles in full; do not abbreviate.

Copies. All manuscripts must be submitted in triplicate. One copy should be the original typed copy, the other two, clear carbon copies or photo reproductions, only. Authors should check the final typing carefully and retain a copy of the manuscript as a precaution against loss in the mail.

Length. Before preparing a manuscript, the author should check several recent issues to get an idea of the approximate length of regular articles published in the *Journal of Applied Psychology*. (One printed page equals roughly three double-spaced typewritten manuscript pages.) A few longer articles of special significance may be printed from time to time as monographs. Occasionally the Journal will have a section of "Short Notes" featuring brief reports on studies which make some methodological contribution or constitute an important replication.

Abstracts. Each copy of the manuscript must be accompanied by an abstract of 100–120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, 16, 833), or they may be obtained from the Editor or from the APA Central Office.

Figures. The original drawing of a figure, or an 8 × 10-in. glossy print of the drawing, is needed for publication and must be submitted with the original typed copy of the manuscript. Duplicate copies of the figure may be photographic or pencil-drawn copies. Figures must be hand-lettered; typewritten lettering is not acceptable.

Early publication. Regular articles are published as nearly as possible in the order in which they were received. Although everything possible is done to keep publication lag to a minimum, at present there is a 10–12-mo. period between the time of receipt and publication of an article. The author of an accepted manuscript may secure early publication, usually in the first issue to go to press after its acceptance, by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication will be sent with the letter of acceptance.

Reprints. No gratis reprints are supplied. Reprints may be ordered from the printer prior to publication.

National Auxiliary Publications Service (NAPS). To save journal space, while at the same time making all necessary materials available to those with a special interest in an article, supplementary materials (such as extensive tables or copies of instruments) may be deposited with the NAPS. Copies of the material to be deposited should be submitted to the Editor with the manuscript and its availability indicated with the following footnote in the article:

> [3] *A five-page table giving data for each of the subgroups* (or other described material) has been deposited with the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

Journal of Applied Psychology
1971, Vol. 55, No. 2, 95–102

Diary No
384
17 · 9 · 71
Lib

THE CULTURE ASSIMILATOR:

AN APPROACH TO CROSS-CULTURAL TRAINING[1]

FRED E. FIEDLER,[2] TERENCE MITCHELL
University of Washington

AND HARRY C. TRIANDIS
University of Illinois

The construction of self-administered programmed culture training manuals, called culture assimilators, is here described. These programs provide an apparently effective method for assisting members of one culture to interact and adjust successfully with members of another culture. Culture assimilators have been constructed for the Arab countries, Iran, Thailand, Central America, and Greece. This article describes the steps involved in the development of these programs, and reviews studies validating the culture assimilator programs under laboratory and field conditions.

Interaction among persons belonging to different culture groups is becoming increasingly common as efforts toward political and economic integration, international cooperation, and technical assistance become more frequent. The literature concerned with stereotyping, interpersonal attitudes, and behavior in culturally heterogenous groups suggests that interaction across cultural barriers produces substantial difficulties (e.g., Fiedler, 1966; Fiedler, Meuwese, & Oonk, 1961; Rombauts, 1962; for a review see Triandis, 1967). A recent handbook of cross-cultural training (Wight, 1969) surveys a large number of approaches which have been proposed to overcome these difficulties. Since most of these approaches have not been subjected to rigorous evaluation, little is known about their effectiveness. One method which has been evaluated and which also seems to produce useful results, is the culture assimilator. The present report is a description and evaluation of this training procedure.

The culture assimilator[3] is a programmed learning experience designed to expose members of one culture to some of the basic concepts, attitudes, role perceptions, customs, and values of another culture. There is no single culture assimilator; different assimilators are needed for each pair of cultures. Given Cultures A, B, and C, we could have assimilators telling A about B, B about A, A about C, C about A, etc. Furthermore, assimilators can be general or specific; that is, they may train an individual to behave effectively in a very narrow set of social situations, such as in a hospital or as a community development worker, or for a broad group of social situations. The content of assimilators can also vary a good deal. Some assimilators emphasize the interpersonal attitudes that contrast the learner's culture and the target culture; other assimilators emphasize the customs of the target culture; and still others concentrate on the value contrasts of the two cultures. Research to date has not progressed sufficiently to indicate which kind of assimilator is maximally useful in which type of social situation. The results this article reviews suggest that a balanced assimilator containing an equal number of general and specific items, and concerned with each of the above areas of content does improve interpersonal interaction in heterocultural settings. More efficient designs for developing these programs are likely to be found in future research.

---

[1] This paper was prepared under Contract N00014–67–A–0103–0013 with the Advanced Research Projects Agency, United States Navy (Fred E. Fiedler, principal investigator). Research reported herein was conducted in part under Contract DA 49–193–MD–2060 with the Office of the Surgeon General, United States Army (Fred E. Friedler, principal investigator) and under ARPA Order 454, Nonr 1834(36) with the Advanced Research Projects Agency and the Office of Naval Research (Fred E. Fiedler & Harry C. Triandis, principal investigators).

[2] Requests for reprints should be sent to Fred E. Fiedler, Department of Psychology, University of Washington, 33 Johnson Hall, Seattle, Washington 98105.

[3] The concept of the culture assimilator training was originated by L. M. Stolurow, now at the State University of New York at Stony Brook, who was one of the co-investigators of the Advanced Research Projects Agency project from 1962 to 1967.

Preparation for overseas work assignments varies from giving a man a two-page pamphlet to several years of preparation. This training will vary depending on the organization, the type of work, and the available time to provide training. Such preparation is a key component in the in-country training of Peace Corps volunteers, while it is very limited in the case of most business organizations. The culture assimilator is designed primarily for personnel with limited time and who, therefore, would normally receive little or no training. It is designed to provide the trainee with extensive information about the culture in a 2–5-hr. time span.

In order to provide such information it is necessary to select those areas of culture contrast which have the largest impact on social behavior across cultural barriers. This requires an exceptionally good understanding of the two cultures. Such understanding can be obtained by examining existing ethnographies and historical records, as well as by analyses of the subjective cultures of the two cultural groups.

By subjective culture we mean the characteristic way in which a cultural group perceives and responds to its social environment. Triandis (Triandis & Vassiliou, 1967; Triandis, Vassiliou, & Nassiakou, 1968; Triandis et al., 1971) and his co-workers have developed a number of methods for the cross-cultural study of perception and cognition. The typical study involves representative samples of the population of each country. Each respondent is asked a large number of questions of the psychological type requiring sentence completions, ratings, comparisons, rank-orderings, the selection of the best response from a list, and so on. Care is taken to ask a representative set of questions concerning a representative sample of social objects. The information is analyzed by means of multivariate methods, particularly factor analysis and analysis of variance, to determine the major dimensions of social perception and cognition used in each culture and the extent to which these dimensions influence a wide variety of responses. An examination of the relationships among such data normally reveals some themes or typical culture-determined viewpoints which can be communicated as representative generalizations about that culture. The assimilator can then focus on those important themes and omit details about the other culture that cannot be easily learned in a short time. However, only in the most recently developed program, the Greek assimilator, was there an extensive use of these methods in the construction of the program, although other programs made indirect use of information gathered by these methods.

The main problem in the development of the culture assimilator, as well as any culture training program, is: How can we meaningfully simulate intercultural interactions? There are innumerable ways of learning about another culture. Most methods involving paper and pencil or books make the trainee a passive observer. Yet the heterocultural encounter is by definition an active process. The participants must continually evaluate, interpret, and react, and again evaluate the effect of their reactions. We were, therefore, committed to an instructional program which made the trainee an active participant in the learning process.

Granted that no book or reading situation will be able to provide ideal simulation experiences, programmed instruction does have the advantage of forcing the individual to interpret and evaluate various situations and to assimilate immediate feedback on the adequacy and accuracy of his interpretation.

Basically, the method here described requires the trainee to read a short episode of an intercultural encounter, and to give an interpretation of the encounter. The program gives an explanation of why his answer was correct or incorrect. If the latter, the trainee is asked to go back to the episode, re-read the material, and choose again. The procedure is thus quite simple, and the overwhelming majority of trainees find it very stimulating and enjoyable. In contrast, the development of an effective and well-written culture assimilator is a rather complex process. The following sections attempt to provide a guide for the construction of these instruments.

STEPS IN THE DEVELOPMENT OF CULTURE ASSIMILATORS

Selection of Content

One basic premise of all culture assimilators is that the critical problems in heterocultural encounters occur in areas in which there are the greatest divergencies in norms, customs, and values between two cultures. Where no differences exist, there will be little cause for conflict. The problem is, thus, one of identifying the culturally critical concepts and behaviors, and in providing the trainee with a series of experiences in which he must symbolically cope with them. A number of methods have been used to identify these critical concepts and behaviors. The analysis of subjective culture, briefly mentioned earlier, is one of these approaches. Another concerns the analysis of critical incidents.

Critical incidents. The critical incident technique was originally developed by Flanagan (1954) who defines it as consisting

> of a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in solving practical problems and developing broad psychological principles. . . . By an incident is meant any observable human activity that is sufficiently complete in itself to permit inferences and predictions to be made about the person performing the act. To be critical, an incident must occur in a situation where the purpose or intent of the act seems fairly clear to the observer and where its consequences are sufficiently definite to leave little doubt concerning its effects [p. 327].

Flanagan (1954) utilized the critical incident technique for the investigation of effective and ineffective job behavior. For the purposes of developing culture assimilators, the ideal incident must describe (*a*) a common occurrence in which an American and a host national interact; (*b*) a situation which the American finds conflictful, puzzling, or which he is likely to misinterpret; and (*c*) a situation which can be interpreted in a fairly unequivocal manner, given sufficient knowledge about the culture. Finally, the incident must be relevant to the American's task or mission requirements. Thus, a love affair may be interesting or upsetting for some Americans overseas, and it may present innumerable occasions which the American finds difficult to understand. However, under ordinary circumstances, information concerning love affairs would not be considered relevant for the job.

Critical incidents are obtained by asking Americans and host nationals with whom they come in contact to describe some specific intercultural occurrences or events that made a major difference in their attitudes or behavior toward the members of the other culture. These may be pleasant, unpleasant, or simply nonunderstandable occurrences. One such incident from the most recently developed assimilator is presented below.

Incident reported by an American teacher in Greece. The teacher mentioned that when she was new in Greece she met a number of Greeks whom she knew only, what she considered to be, casually. She was surprised by the fact that these people often asked her questions which were of a rather personal nature. She felt that the Greeks were somewhat rude to ask questions about her clothes, standard of living, and other personal things. She felt that this information should be told only to close friends.

The incident is then refined into the episode which, along with 75–100 others, is eventually used in the program. Names are changed and irrelevant material dropped. An example of the final episode is below:[4]

Page X–1

Sharon Hatfield, a school teacher in Athens, was amazed at the questions that were asked her by Greeks whom she considered to be only casual acquaintances. When she entered or left her apartment, people would ask her where she was going or where she had been. If she stopped to talk she was asked questions like, "How much do you make a month?" or "Where did you get that dress you are wearing?" She thought the Greeks were very rude.

Alternatives and feedback. Having written the episodes, it is then necessary to write four possible responses which will be presented to the trainee. Each of these four alternatives obviously must be plausible. In fact, the preferred method is to ask the Americans interviewed to write out various explanations, and the most plausible are selected from this pool of alternatives. Only one is deemed correct, with the others containing various ethnocentric errors.

---

[4] Each set of statements here shown in a box is normally presented on a separate page.

Finally, four feedback explanations must be written. These tell the trainee not only whether his answer is correct or incorrect, but they also provide him with further information about the culture. They should, in effect, give the trainee a rationale for interpreting the correctness or incorrectness of his reply and assist him in building up a frame of reference for handling similar situations. The feedback information is taken from our interviews and subjective culture data. Examples of the alternatives and feedback explanations are below:

Page X-2

Why did the Greeks ask Sharon such "personal" questions?

1. The casual acquaintances were acting like friends do in Greece, although Sharon did not realize it.

   Go to page X-3

2. The Greeks asked Sharon the questions in order to determine whether she belonged to the Greek Orthodox Church.

   Go to page X-4

3. The Greeks were unhappy about the way in which she lived and they were trying to get Sharon to change her habits.

   Go to page X-5

4. In Greece such questions are perfectly proper when asked of women, but improper when asked of men.

   Go to page X-6

Page X-3

You selected 1: The casual acquaintances were acting like friends do in Greece, although Sharon did not realize it.

Correct. It is not improper for in-group members to ask these questions of one another. Furthermore, these questions reflect the fact that friendships (even "casual" ones) tend to be more intimate in Greece than in America. As a result, friends are generally free to ask questions which would seem too personal in America.

Go to page X-1

Page X-4

You selected 2: The Greeks asked Sharon the question in order to determine whether or not she belonged to the Greek Orthodox Church.

No. This is not why the Greeks asked Sharon such questions. Remember, whether or not some information is "personal" depends upon the culture. In this case, the Greeks did not consider these questions too "personal." Why? Try again.

Go to page X-1

Page X-5

You selected 3: The Greeks were unhappy about the way in which she lived and they were trying to get Sharon to change her habits.

No. There was no information given to lead you to believe that the Greeks were unhappy with Sharon's way of living. The episode states that the Greeks were acquaintances of Sharon.

Go to page X-1

Page X-6

You selected 4: In Greece such questions are perfectly proper when asked of women, but improper when asked of men.

No. Such questions are indeed proper under certain situations. However, sex has nothing to do with it. When are these questions proper? Try to apply what you have learned about proper behavior between friends in Greece. Was Sharon regarded as a friend by these Greeks?

Go to page X-1

One of the most important functions of feedback is to expose the learner to some of the major themes characterizing the two cultures with which he is working. In short, information learned from the analyses of the subjective cultures of the two communities can be incorporated in the feedback. New concepts can thus be taught in the context of concrete incidents.

Validation Steps

After incidents have been constructed and the alternatives and feedback are prepared, a number of validation steps are undertaken. All the stories with just the second page of alternatives (i.e., without the feedback) are submitted to an appropriate sample of people from the target culture. Each subject is asked to read the story and to choose the alternative which he feels is most applicable as a description of what occurred from his culture's point of view. This information is then tallied and only those stories for which there is a high percentage of agreement are incorporated into the final product. The program is, therefore, based on a model which holds that given behaviors have meanings which are more probable than others. Most of the time when a certain situation or behavior occurs, the person from the other culture means a certain thing by his action.

Members of the same subject population are also asked to rate each story on how important they feel the episode is, and how probable is its occurrence. The results from these scales help us further in deciding which incidents would represent the most frequent and important problems that might occur.

When the final group of incidents has been chosen, they are examined by a number of judges and sorted into categories according to the cultural concept with which they deal—for example, hospitality, authority relations, saving face, etc. These groups of incidents are then sequenced according to two general principles: (*a*) within a category (i.e., incidents dealing with one concept or custom) the episodes begin with simple items and progress to more difficult ones (see Foa & Chemers, 1967); (*b*) across categories of concepts the program generally deals with rather broad social issues at first and then moves to more specific problems. Introductory and summary statements (one or two paragraphs) are presented before and after large groups of incidents dealing with similar topics, and all of these incidents and statements are then bound together in book form.

Depending in part on the availability of critical incidents, we have generally obtained 150–200 incidents, of which 75–100 are eventually included in the culture assimilator. The time for preparing a 75–100-item program is estimated as approximately 800 man-hours spread over 9 mo. This includes collection of items, writing the episodes and alternatives, and checking and revising the alternatives and feedback information.

RESULTS FROM LABORATORY EXPERIMENTS

Four studies have been completed which provide validation evidence for the culture assimilator programs. These studies are very briefly summarized below.

Arab Assimilator

A laboratory study (Chemers, Lekhyananda, Fiedler, & Stolurow, 1966) tested the Arab culture assimilator which was the first program and, therefore, relatively primitive. The *S*s consisted of 27 American Reserve Officer Training Corps (ROTC) cadets who served as leaders. In addition, 14 male members of the University of Illinois Arab Student Association participated as paid group members. These Arab students worked in two-man teams with various American leaders to whom they were assigned at random.

The group of leaders (ROTC cadets) were randomly dichotomized into a control group which received programmed instruction on Arab country geography. The other 14 leaders received the culture assimilator which dealt with the role of women, importance of religion in the Mid East, and interpersonal skills in group interaction. The testing and training sessions took place on four consecutive nights.

After the training period, each American was appointed the leader of a three-man group with two Arabs as members. Each group performed three tasks: (*a*) a letter encouraging Arab villagers to allow their women to work in a factory to be built nearby; (*b*) traversing a road map with the shortest possible mileage, given that every point on the map had to be touched at least once; and (*c*) a negotiation task to decide on the number of Moslems to be employed in an Arab-American mining venture. The Americans were told to seek a nondiscriminatory hiring policy. The tasks were administered in counterbalanced order, and the pairs were systematically rotated to balance the design of the study. After each task all participants completed various scales describing the group climate, perceived effectiveness, stressfulness of the situation, etc.

The differences in task performance between culture assimilator-trained leaders and leaders with geography training were small, but in the hypothesized direction, although none reached an acceptable significance level. These evaluations consisted of two ratings by judges for the first task; time to complete the course for the second task; and seven scores composed of ratings and ranking of the negotiated settlements for the third task. Also, 9 of the 11 indexes of leader-member relations were in the expected direction with one at the $p < .05$ level, and one at the $p < .10$ level. These 11 indexes consisted of three measures generated by leader ratings; five measures composed of member ratings, and three combined ratings of the overall group atmosphere. The results of this pilot study, therefore, yielded promising relations which encouraged further work.

Thai Culture Assimilator

A second study used a considerably more sophisticated culture assimilator developed for Thailand (Foa, Mitchell, Santhai, Wichiarajote, & Wichiarajote, 1967). This Thai culture assimilator was experimentally tested by Mitchell and Foa (1969). The study involved 32 American ROTC cadets of a Special Forces company who were matched according to rank, and then randomly assigned to an experimental and a control group. The 16 men of the experimental group received the culture assimilator program, and the 16 men of the control group received the geography program.

Sixteen pairs of foreign students from the Far East participated in the study. One member of each pair was from Thailand, the other from Korea or another Far Eastern country. Each pair of foreign students worked with one culture-trained and one geography-trained American. The American's job was to supervise the construction of a toy building made by his team of two foreign students and to get the building constructed in the shortest possible time and as accurately as possible. The American leader was not, however, permitted to work on the task. The experimental situation was conceptualized as simulating the problems of a Western expert working in a developing country: the expert is usually better qualified technically than his local co-workers and he is expected to train them in doing the job rather than doing it himself.

At the end of each session, a Thai observer, the American leader, and the two foreign students completed an evaluation form. After the second session, all but the leader answered an additional rating form comparing the two American leaders. Since culture- and geography-trained Americans alternated in the sessions, this last rating amounted to a comparison of the two types of training, although neither the group participants nor the Thai observer knew that they were making this type of comparison. Since there were two observers and two task orders possible, the design was counterbalanced to control for these possible confounds.

Three composite ratings of the interpersonal relations were generated for each of the four participants. These ratings dealt with the leader's handling of the group as a whole, his treatment of the members, and an overall assessment of the group climate. For these three ratings, two were significant for the leader ratings, all three for the Thai observer, one for the Thai member, and none for the Korean member. All 12 ratings were in the predicted direction.

In the area of group effectiveness, two scale ratings were used as criteria as well as measures of the number of mistakes made in the construction of the toy buildings and the time to finish the task. None of these ratings or measures were significant. The results were, therefore, similar to those of the first study showing that the interpersonal area seems to be more sensitive to the culture training than group effectiveness. Whether these interpersonal differences would facilitate productivity over a more prolonged period of time in a field situation was, of course, a question for subsequent field studies.

FIELD STUDIES

Honduras Culture Assimilator

The third study involved a program for Central America (Symonds, O'Brien, Vidmar,

& Hornik, 1967) which was experimentally validated in a study by O'Brien, Fiedler, and Hewett (1970). This study was conducted in the context of a people-to-people program conducted by Los Amigos de las Americas of Houston, Texas. Amigos yearly sends 250–300 teenagers to Central America to establish and operate public health clinics which provide polio vaccinations, diptheria shots, and similar preventive measures, as well as performing community development work. Three waves of teen-aged participants went to Honduras and Guatemala for 3-wk. terms. In their host country the members of each wave were assigned to teams of 2–6 persons, as well as to a national and three regional headquarters teams.

Of the 265 participants in the summer of 1967, 119 received the culture assimilator training while the remaining 146 received the culture and area training given normally. To minimize the possibility of a Hawthorne effect the culture assimilator training was given no particular emphasis. The training given to Amigos varied somewhat; that is, some Amigos heard some lecturers or saw some films and not others, depending upon the available time and the exigencies of scheduling. The culture assimilator training was, therefore, not seen as anything out of the ordinary.

The culture assimilator for Honduras and Guatemala was based on critical incidents reported by Amigos who had participated in the program the previous year. The program used a theoretical framework of structural role theory (Oeser & Harary, 1962, 1964; Oeser & O'Brien, 1967) and emphasized incidents dealing with relations between the Amigos and host country nationals collected from the previous year's group.

The evaluation of performance and adjustment took two forms: (*a*) performance ratings were obtained from the director of the Amigos program and from his various headquarters staff members; (*b*) in addition, several adjustment measures were obtained from the Amigos themselves. These included (*a*) self-ideal—self-discrepancy scores obtained before and after the overseas experience, and (*b*) daily ratings of adjustment, mood and satisfaction with self, fellow team members, and work overseas (these came to be known as happy sheets). These latter comparisons were made between first- and third-week ratings.

The results showed a general improvement in work performance of culture assimilator-trained as against untrained Amigos. Of particular interest was the finding that the mean adjustment of all trained Amigos also generally improved, but that the improvement as well as the work performance was higher for those who had been Amigos the previous year than for those who had gone to Central America for the first time. This completely unexpected finding suggests that the culture assimilator effectively assists in the integration of previously obtained cultural experiences as well as providing information for the novice abroad.[5] The data thus show that the assimilator training contributed significantly to adjustment as well as performance in field conditions and to better interpersonal relations in laboratory studies.

Thai Assimilator Field Study

Finally, Worchel and Mitchell (1970) recently completed a field test with the Thai Assimilator, utilizing as *S*s officers of the United States Military Advising and Assistance Group and civilians from the United States. The available officers and civilians were randomly divided into those who received the culture assimilator training and those who received comparable culture training in essay form.

The method of evaluating the training consisted of self-report questionnaires mailed out several months after the training. Other evaluations, while highly desirable, could not be obtained on advice of the American authorities in Thailand.

The questionnaire contained 15 items about the officer's experience in Thailand. These items dealt with his contacts with Thais, his work relationship, his health, his desire to go back to Thailand, etc. On 12 of these items, the differences between respondents with culture assimilator training were more favorable, and of those, 4 were statistically significant. None of the items in the counter-expectational direction approached significance. Thus, the culture assimilator training could be regarded as successful.

---

[5] This finding is supported by informal statements of personnel of the United States mission in Iran where an experimental study, using an Iranian culture assimilator, was conducted by Chemers (1968). Unfortunately, the Iran assimilator has not been tested under field conditions as yet.

Greek Assimilator Field Study

A recently completed Greek culture assimilator (Mitchell, Gagerman, & Schwartz, 1969) was administered to 76 American military and civilian advisors in Greece. A comparison of culture assimilator-trained and control *S*s showed that the assimilator-trained *S*s reported themselves to be significantly better in getting along with Greeks, in obtaining cooperation from Greeks, in being accepted by Greeks, in understanding Greeks, as well as enjoying their tour of duty and feeling generally better during their tour of duty than did a large group of military personnel who had not received culture assimilator training. The assimilator-trained *S*s also saw themselves as significantly better adjusted and as having significantly better interpersonal relations and higher productivity than *S*s who had received no training. A more detailed report of this study appears in Worchel and Mitchell (1970).

In summary, the major steps in the construction of culture assimilators were described. They provide an apparently useful method for the instruction of members of one culture to behave effectively in social and task situations involving members of another culture. Three particular strengths of this approach are suggested: (*a*) the training depends on a detailed analysis of empirically derived critical cultural information relevant for effective interaction; (*b*) it provides an active, ego-involving training procedure; and (*c*) it has been shown to improve the effectiveness and satisfaction of those trained with this method when they are compared with those trained with other methods.

REFERENCES

Chemers, M. M. *Cross-cultural training as a means for improving situational favorableness.* Urbana, Illinois: Group Effectiveness Research Laboratory, University of Illinois, 1968.

Chemers, M. M., Lekhyananda, D., Fiedler, F. E., & Stolurow, L. M. Some effects of cultural training on leadership in heterocultural task groups. *International Journal of Psychology*, 1966, **1**, 301–314.

Fiedler, F. E. The effect of leadership and cultural heterogeneity on group performance: A test of the contingency model. *Journal of Experimental Social Psychology*, 1966, **2**, 237–264.

Fiedler, F. E., Meuwese, W. A. T., & Oonk, S. Performance of laboratory tasks requiring group creativity. *Acta Psychologica* 1961, **18**, 100–119.

Flanagan, J. C. The critical incident technique. *Psychological Bulletin*, 1954, **51**, 327–358.

Foa, U. G., & Chemers, M. M. The significance of role behavior differentiation for cross-cultural interaction training. *International Journal of Psychology*, 1967, **2**, 45–58.

Foa, U. G., Mitchell, T. R., Santhai, S., Wichiarajote, N., & Wichiarajote, W. *Thai culture assimilator.* Urbana, Illinois: Group Effectiveness Research Laboratory, University of Illinois, 1967.

Mitchell, T. R., Gagerman, J., & Schwartz, S. *Greek culture assimilator.* Urbana, Illinois: Group Effectiveness Research Laboratory, University of Illinois, 1967.

Mitchell, T. R., & Foa, U. G. Diffusion of the effect of cultural training of the leader in the structure of heterocultural task groups. *Australian Journal of Psychology*, 1969, **21**, 31–43.

Oeser, O. A., & Harary, F. A mathematical model for structural role theory I. *Human Relations*, 1962, **15**, 89–109.

Oeser, O. A., & Harary, F. A mathematical model for structural role theory II. *Human Relations*, 1964, **17**, 3–17.

Oeser, O. A., & O'Brien, G. E. A mathematical model for structural role theory III. *Human Relations*, 1967, **20**, 83–97.

O'Brien, G. E., Fiedler, F. E., & Hewlett, T. *The effects of programmed culture training upon the performance of volunteer medical teams in Central America.* Seattle: Organizational Research, University of Washington, 1970.

Rombauts, J. Gedrag en groepsbeleving in etnischhomogene en etnischheterogene groepen. Unpublished doctoral dissertation, Louvain, Belgium: University of Louvain, 1962.

Symonds, J., O'Brien, G., Vidmar, M., & Hornik, J. *Honduras culture assimilator.* Urbana, Illinois: Group Effectiveness Research Laboratory, University of Illinois, 1967.

Triandis, H. C. Interpersonal relations in international organizations. *Organizational Behavior and Human Performance*, 1967, **2**, 26–55.

Triandis, H. C., & Vassiliou, V. Frequency of contact and stereotyping. *Journal of Personality and Social Psychology*, 1967, **7**, 316–328.

Triandis, H. C., Vassiliou, V., & Nassiakou, M. Three cross-cultural studies of subjective culture. *Journal of Personality and Social Psychology*, 1968, **8**, (4, Pt. 2).

Triandis, H. C., Vassiliou, V., Vassiliou, G., Tanaka, Y., & Sharmugan, A. V. *The analysis of subjective culture.* New York: Wiley, 1971.

Wight, A. R. *A draft handbook for cross-cultural and community involvement training.* Estes Park, Colorado: Center for Research and Education, 1969.

Worchel, S., & Mitchell, T. R. *An evaluation of the effectiveness of the culture assimilator in Thailand and Greece.* Seattle: Organizational Research, University of Washington, 1970.

(Received February 6, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 103–109

UTILITY OF MERIT INCREASES[1]

BRIAN A. GILES AND GERALD V. BARRETT[2]

Management Research Center, College of Business Administration, University of Rochester

A magnitude estimation technique was used to determine the relationship between merit increases and satisfaction for professional employees of an electronics organization. The linear, logarithmic, and power functions were fitted to the data in combination with four frames of reference. The power function and the percentage of perceived equitable merit increase frame of reference gave the best fit to the data. The generally accepted Bernoullian concept of diminishing utility for each added dollar was not supported. The analyses found that each dollar of merit increase had increased utility. For approximately 10% of the sample, satisfaction was not a monotonic increasing function of dollars. This is contrary to accepted economic theory but is consistent with equity theory.

There is a long history in theorizing about the subjective value of money to individuals. As far back as 1728, the mathematician Cramer (Bernoulli, 1964) postulated that the value of money might be a power function of the number of dollars. A short time later, in 1738, the philosopher Bernoulli hypothesized a logarithmic function. The importance of this theory has been recognized by republication in modern day literature (Bernoulli, 1964). More specifically, Bernoulli postulated that an individual's wealth influenced the utility of additional money so that an additional dollar has more utility for the poor than the rich man. Since that time, the economic principle of decreasing marginal utility, that each added dollar has less utility (value) than the preceding one, has been widely accepted (Stevens, 1966). The one exception to this principle would be the Friedman and Savage (1948) proposition that increasing marginal utility may come about when a consumer is shifted into a new class that would give him new social and economic status.

While Bernoulli's hypothesis has become well known (Adams, 1963), Cramer's proposition has passed into obscurity and only recently has been revitalized by the extensive work of Stevens. Using a technique of magnitude estimation, Stevens (1966) has consistently shown that a power function best describes the relations between stimulus and response for a number of sensory modalities. More recently, the technique has been applied to attitudes and other social phenomena. Galanter (1962) reported an investigation in which *S*s were asked how much money would make them twice as happy as \$10, \$100, and \$1,000. The results supported the general power law with an exponent of approximately .43 and, at the same time, the proposition of decreasing marginal utility. In a more recent study of crime, Sellin and Wolfgang (1964) found that the seriousness of an offense which involved money grew as a power function of the amount of money stolen. The obtained exponent was .17 which supports even more strongly the proposition of declining marginal utility of money.

In critiquing Galanter's (1962) study as preliminarily reported by Stevens (1959), Edwards (1961) indicates that much more evidence would be required to convince researchers that the utility of money was not linear. In essence, Edwards contends that twice as much money should be twice as important to the individual.

There is one common element among the theoretical propositions of Bernoulli (1964), Cramer (Bernoulli, 1964), Stevens (1959), and Edwards (1961). All would postulate that the utility of money is a monotonic increasing function of the number of dollars. They differ in that Bernoulli would postulate a logarithmic function, Cramer and Stevens a power func-

---

[1] This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Gerald V. Barrett, Management Research Center, College of Business Administration, University of Rochester, Dewey, Rochester, New York 14627.

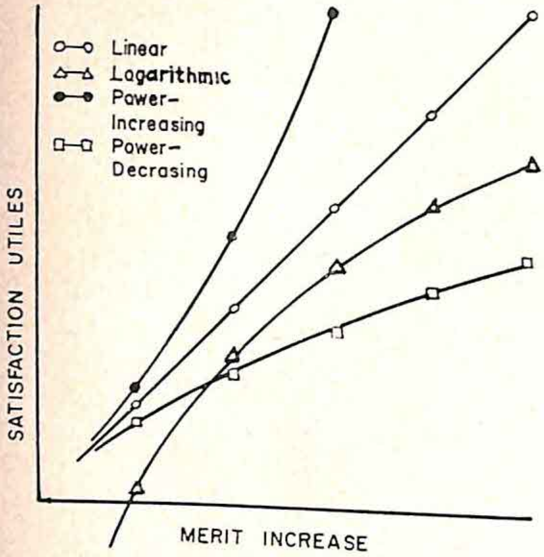


FIG. 1. Representation of linear, logarithmic, and power functions. (A utile is a hypothetical unit of utility.)

tion, and Edwards a linear function. The various theoretical curves are illustrated in Figure 1. If the exponent of the power curve is less than 1, this represents decreasing marginal utility, and if more than 1, increasing marginal utility.

While there seems to have been a great deal of speculation about the utility of money, especially in economics, there seems to be very little empirical work concerning the actual utility of money. No reports were found which studied the utility of merit increases. The closest topic was a recent study by Zedeck and Smith (1968) which used the psychophysical method of limits to determine thresholds of perceived equitable payment. Their concern was with determining what was considered to be a fair salary. The present investigation's focus was on perceived equitable merit increases and the utility (satisfaction) associated with increases above and below that point.

Specifically, the purpose of the present investigation was to determine if the logarithmic, linear, or power function best fits the relationship between specified merit increases and utility (individual satisfaction), and to determine the empirical relationship using the appropriate model. The factors of profession, age, and years of service were examined for moderating effects on the relationship.

METHOD

Subjects

Questionnaires were distributed to 157 professional level personnel employed at a medium-sized Eastern electronics company. Eighty *Ss* responded to the anonymous questionnaire. The 64 replies (41% of the population) containing complete data, were used in the present investigation.

The sample included 31 professionals classified as engineers and 33 classified as nonengineers who were in the fields of accounting, purchasing, production, marketing, and personnel. Thirty-nine *Ss* had a bachelors degree or beyond while 17 had completed 2 yr. at college. The remaining eight *Ss* had a high school education. Fifty-three *Ss* were between the ages of 25 and 39. Only five *Ss* were younger than 24, and six were over 45. A total of 29 *Ss* reported their yearly salary as being between \$8,000 and \$12,000 a year. Nine *Ss* each reported income in the brackets of \$12,000–\$14,000, \$14,000–\$16,000, and \$16,000–\$18,000. Eight *Ss* reported an annual salary of over \$18,000 a year.

Research Instrument

The questionnaire used for the present study was similar in its basic construction to one that has been previously used in investigations of juvenile delinquency (Sellin & Wolfgang, 1964) and information systems (Barrett, Thornton, & Cabe, 1968). The first page of the questionnaire was as follows:

> Most companies award merit increases to salaried employees as part of an annual performance appraisal. No doubt you look forward to your next performance appraisal with expectations of receiving a certain increase.
>
> Merit increases in dollars per year
>
> \$1,700 ☐
> 500 ☐
> 1,000 ☐
> 900 ☐
> 2,100 ☐
> 600 ☐
> 1,100 ☐
>
> Instructions
>
> 1. Fourteen possible merit increases are listed to the left. Choose and write on the line below the merit increase which you feel would be fair and equitable at your next yearly performance appraisal based on your own perceived job performance and contributions to the company.
>
> ________ 10
>
> 2. Your next task is to show your satisfaction with each of the remaining merit increases. You do this by writing down in the score box beside each merit increase a number which shows your satisfaction in relation to the fair and equitable merit increase chosen above.
>
> Your fair and equitable merit increase has been assigned a value of ten. Use this merit increase as a standard and score every other merit increase in relation to this standard. For example, if any

other merit increase would seem to bring you five times as much satisfaction, write in a score of 50. If a merit increase would bring you only half as much satisfaction, write in a score of 5.0. If a merit increase would seem to bring only 1/20 of the satisfaction as the standard, write in a score of 0.5.

You may use any whole or decimal numbers that are greater than zero, no matter how small or how large they are, just so long as they represent how much satisfaction the merit increase would bring you as compared to the standard merit increase.

700 ☐
2,500 ☐
1,300 ☐
800 ☐
300 ☐
1,500 ☐
3,000 ☐

The second page of the questionnaire elicited information concerning *S*'s annual salary, length of service with the company, occupation, age, number of dependents, and formal education.

Procedure

Each professional employee in the company was sent an introductory letter explaining the purpose of the investigation. They were asked to fill out the enclosed questionnaire and to return the questionnaire unsigned in an envelope to an address outside the company. Two weeks after the initial letter, a follow-up letter was sent to all professional employees thanking those who had returned the questionnaire and requesting those who had not to please return it.

RESULTS

Twelve separate regression and correlation analyses were performed on the data taking into account four frames of reference and three hypothesized functions as shown in Table 1. In each case, there were 64 *S*s making 14 judgments resulting in 896 data points. The four frames of reference or predictors were chosen on either a logical or theoretical basis. It was assumed that the combination frame of reference and function which resulted in the highest correlation would best represent *S*s decision process. The simplest frame of reference chosen was the correlation of each of the 14 merit increases with the satisfaction values ascribed to them. The second frame of reference was the percentage merit increase, obtained by dividing each of the 14 merit increases by each *S*'s salary which was correlated with the satisfaction values for each merit increase. The third frame of reference was the percentage of which each of the 14 merit increases represented perceived equitable merit increase for each respondent. This was considered to be the frame of reference most analogous to the questionnaire instructions which requested that each employee compare his chosen equitable merit increase with every other increase. The last predictor or frame of reference was most analogous to Bernoulli's (1964) concept of wealth. The employee's salary plus merit increase was considered to be the best estimate of his total wealth. In Table 1, U is the utility or perceived satisfaction, M is the merit raise, S is current salary, and E is the perceived equitable merit increase.

The simplest frame of reference or independent variable was the dollar value given to each of the 14 merit increases. As shown in Table 1 the merit increase frame of reference gave the highest correlation with the power function, but this coefficient of .76 was not significantly different from the .73 obtained with the linear function. Both the linear and power correlations were significantly higher ($p < .05$; z transformation) than the .66 correlation obtained with the logarithmic function.

The second frame of reference or independent variable was the percentage merit increase. As shown in Table 1 the correlations were almost identical to the merit increase frame of reference. The third frame of reference was the percentage of perceived equitable merit increase. The .79 correlation obtained with the power function was significantly higher ($p < .05$) than either the linear or logarithmic function. The last frame of reference was wealth as a predictor of satisfaction. As shown in Table 1 all the obtained correlations were low.

The results indicate that the most consistent relationship was found with the power function over all frames of reference. The highest correlation was obtained with the percentage of perceived equitable merit increase frame of reference expressed as a power function. This obtained regression line is shown in Figure 2.

To determine if the factors of age, profession, and service generated significantly different relationships between merit increases and satisfaction, separate correlation and regression analyses were computed using the percentage of perceived equitable merit increase frame of reference and the power function which had given the highest correlation.

As shown in Table 2, the only significant difference in these divisions was between engineers and nonengineers ($p < .05$).

A surprising finding was that seven of the employees did not report satisfaction as a monotonically increasing function of merit increase. For those seven employees, more money did not bring more satisfaction, and for some, more money would actually result in a decrease in utility. Despite this, the correlations reported (see Table 1) were not significantly different from those obtained when the group was removed from the total sample.

DISCUSSION

Overall, the results gave the most support to the power function and least support to the logarithmic function. The percentage of perceived equitable merit increase frame of reference and the power function best represented the data. The concept of wealth (current salary plus the merit increase) did not seem to be a predictor of importance considering the negligible correlations obtained when it was used with all three functions.

The failure of support for Bernoulli's (1964) decreasing marginal utility of money is consistent with other studies. Allais (1953) presented rational arguments against Bernoulli's concept while an experimental study by Coombs and Pruitt (1960) failed to support the concept. The complexity and confounding influence of the experimental population was illustrated by Mosteller and Nogee (1951) who found that, with the college student *S*s, support was given to the diminishing utility of money while with national guardsmen, the data indicated increasing utility of money. A recent financial portfolio simulation by Gordon, Paradis, and Rorke (1969) found the power function to best represent the utility of money, but the coefficient was below one supporting decreasing marginal utility.

One of the most intriguing supports for both the power function and the increasing marginal utility of money was presented by Starr (1969) who compared miner's hourly wages in coal, metal, and nonmetal mining with the accident and severity rate of mining injuries in these fields of mining. A power function was obtained with an exponent of approximately 3. This result can be interpreted as support for the increasing marginal

TABLE 1

CORRELATION (PRODUCT–MOMENT) BETWEEN SATISFACTION AND MERIT INCREASE WITH FOUR FRAMES OF REFERENCE AND THREE DIFFERENT FUNCTIONS

Frame of reference	Function		
	Linear	Logarithmic	Power
Merit increase (M)	$U = A + B\ (M)$ .73	$U = A + B \operatorname{Log}\ (M)$ .66	$U = A + B\ (M)^n$ .76[a]
Percentage merit increase (M/S)	$U = A + B\ (M/S)$ .73	$U = A + B \operatorname{Log}\ (M/S)$ .67	$U = A + B\ (M/S)^n$ .76[b]
Percentage of perceived equitable merit increase (M/E)	$U = A + B\ (M/E)$ .23	$U = A + B \operatorname{Log}\ (M/E)$ .68	$U = A + B\ (M/E)^n$ .79[c]
Wealth ($S + M$)	$U = A + B\ (S + M)$ .00	$U = A + B \operatorname{Log}\ (S + M)$ .17	$U = A + B\ (S + M)^n$ .21[d]

Note.—Abbreviations: U = utility or perceived satisfaction, M = merit raise, S = current salary, E = perceived equitable merit increase.

[a] $n = 2.26$.
[b] $n = 2.06$.
[c] $n = 2.21$.
[d] $n = .16$.

utility of money since miners appear voluntarily to work in situations involving more risk for each additional dollar they are paid. Of course, an alternate interpretation is that the risk is not perceived according to objective reality, but the fact remains society pays individuals in the occupation as if money had increasing marginal utility.

The widely cited Friedman and Savage (1948) curve of utility does embrace both the concept of increasing and decreasing marginal utility, but suffers from a number of defects from the viewpoint of an experimentalist attempting to test the theory. As Alchian (1953) has pointed out, it is not clear whether an individual moves up or down the curve as salary increases or whether current income is a reference point somewhere on the curve for changes in income. While it is postulated that increasing marginal utility occurs when an individual moves into a new economic class, what constitutes a new class is not defined. The theory's generality makes definitive test almost impossible. In our judgment the Friedman and Savage model does not adequately explain the results of this and other empirical studies. The failure to be able to test empirically specific features of the model may account for its survival for over 20 yr. While it is impossible to make a final statement at this time, based on this and other studies, strong support can be given for utility being a power function of money with the increasing or decreasing marginal utility being a function of both the situation and the characteristics of *Ss*.

TABLE 2

CORRELATIONS (PRODUCT–MOMENT) BETWEEN PERCENTAGE EQUITABLE MERIT INCREASE AS A POWER FUNCTION OF SATISFACTION WITH THE SAMPLE DIVIDED BY AGE, PROFESSION, AND YEARS SERVICE IN COMPANY

Sample	N	Correlation	n
Age			
Under 30	24	.79	2.28
30 and over	40	.80	2.17
Profession			
Engineer	31	.77*	2.16
Nonengineer	33	.82*	2.27
Service			
Under 3 yr.	35	.78	2.15
3 yr. and over	29	.81	2.29

* $p = .05$.

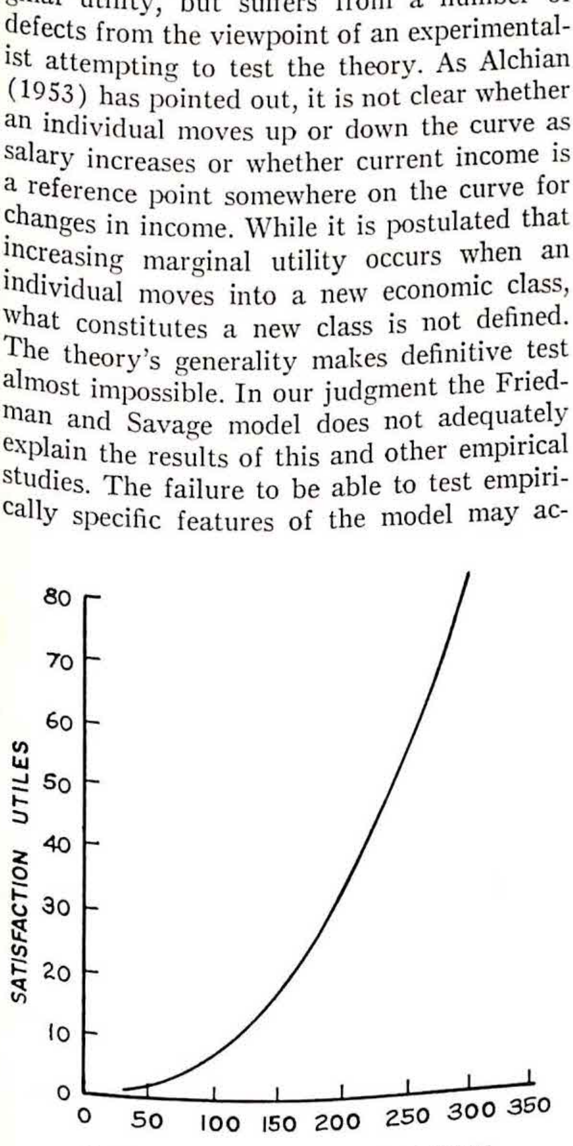


FIG. 2. Obtained regression line for percentage of perceived equitable merit increase as a predictor of utility. (A utile is a hypothetical unit of utility.)

It would appear that our results from the equity group (seven employees for whom utility was not a monotonically increasing function of money) are another example of utility function confounding (Conrath & Deci, 1969). It has been found, especially in game situations, that occasionally the utility for outcomes is not monotonically related to the rewards ascribed to the outcomes. This phenomena often occurs in an interpersonal interaction where not only the individual's absolute reward, but reward relative to others, is involved in the choice process. The utility function confounding obtained in the present investigation appears not to be a result of comparison with someone else, but an estimation of equity and relative risk. From the comments received by slightly over 10% of the sample whose satisfaction was not a monotonically increasing function of salary, there was a point reached where a larger merit increase gave cause for worry. The worry expressed was related both to the feeling that a very large increase in salary would be unearned and that there would be performance expectations which could not be met. From the comments received, it is evident that a small minority of employees would not be happy with extremely large increases in salary.

There appears to be a mixture of both equity and risk involved in the perception of those whose utility functions are not monotonically increasing. The equity component relates to their desire not to be overpaid for the responsibility of their job, while some perceive a risk in a large salary increase since this might imply an expectation of higher job performance than the individual could meet. The perceived inequity of too large a raise seems to fit nicely Adams' (1963) theory of equity.

The company investigated was a relatively young, fast growing, technically based organization. We could speculate that a company with different characteristics might have employees with a different power function and percentage in the equity group. For example, one would predict that an older, more conservative firm would have a larger percentage in the equity group and a smaller exponent in the power function. It could be hypothesized that in a more conservative organization which gave smaller salary increases, a large increase would evoke more feelings of inequity and high risk. In the same fashion, a more conservative firm might place more emphasis on security rather than pay as a reward resulting in less satisfaction with pay. This technique might conceivably provide an objective basis for differentiating organizations.

The differences in individual responses to merit increases suggest that personality variables may be important determinants, especially when comparing the equity group with the other respondents.

In opposition to the generally accepted principle of diminishing utility of each added dollar, the present investigation found that each additional dollar had increased utility. In addition, a small minority of the employees had increasing utility up to a point and then an additional dollar gave not diminishing, but the same or less utility than the previous dollar. Traditional economic theory appears to be unable to embrace these findings. If these results are replicated, it would appear that conventional economic concepts must take into account other complex psychological variables which elicit differential responses from employees.

It is recognized that the magnitude estimation technique does have certain shortcomings. Most critical perhaps for our present investigation was the number 10 used as a standard. It is known that size of the number influences the exponent in the power function (Poulton, 1968). In addition, both the range and distribution of merit increases used in the questionnaire may influence the response.

Some *S*s in responding to a questionnaire of this type might attempt to order the merit increases in ascending values and adjust their responses accordingly. For this reason, the increments between listed merit increases were not uniform. The bunching of merit increases between \$500 and \$1,100 also served the purpose of collecting close incremental responses above and below the values most likely to be chosen as an equitable merit increase. This nonuniform spacing or bunching however, may in itself, influence the consistency of *S*s' response.

One solution to this possible tendency of *S*s to compare and systematize their responses would be to present each merit increase on a separate sheet of paper or a projector. Future investigations will have to take into account these methodological considerations.

REFERENCES

Adams, J. S. Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 1963, **67**, 422–436.

Alchian, A. The meaning of utility measurement. *American Economic Review*, 1953, **43**, 26–50.

Allais, M. Le comportement de l'homme rationnel devant le risque: Critiques des postulats et axiomes de l'ecole americaine. *Econometrica*, 1953, **21**, 503–546.

Barrett, G. V., Thornton, C. L., & Cabe, P. A. Human factors evaluation of a computer based information storage and retrieval system. *Human Factors*, 1968, **10**, 431–436.

Bernoulli, D. Exposition of a new theory on the measurement of risk. In G. A. Miller (Ed.), *Mathematics and psychology*. New York: Wiley, 1964.

Conrath, D. W., & Deci, E. L. The determination and scaling of a bivariate utility function. *Behavioral Science,* 1969, **14**, 316–327.

Coombs, C. H., & Pruitt, D. G. Components of risk in decision making: Probability and variance preferences. *Journal of Experimental Psychology,* 1960, **60**, 265–277.

Edwards, W. Behavioral decision theory. *Annual Review of Psychology,* 1961, **12**, 473–498.

Friedman, M., & Savage, L. J. The utility analysis of choices involving risk. *The Journal of Political Economy,* 1948, **56**, 279–304.

Galanter, E. The direct measurement of utility and subjective probability. *American Journal of Psychology,* 1962, **75**, 208–220.

Gordon, M. J., Paradis, G. E., & Rorke, C. H. An experimental model of portfolio behavior. Unpublished manuscript, College of Business Administration, University of Rochester, 1969.

Mosteller, F., & Nogee, P. An experimental measurement of utility. *The Journal of Political Economy,* 1951, **59**, 371–404.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin,* 1968, **69**, 1–19.

Sellin, T., & Wolfgang, M. E. *The measurement of delinquency.* New York: Wiley, 1964.

Starr, C. Social benefit versus technological risk. *Science,* 1969, **165**, 1232–1238.

Stevens, S. S. Measurement, psychophysics, and utility. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories.* New York: Wiley, 1959.

Stevens, S. S. A metric for the social consensus. *Science,* 1966, **151**, 530–541.

Zedeck, W., & Smith, P. C. A psychophysical determination of equitable payment: A methodological study. *Journal of Applied Psychology,* 1968, **52**, 343–347.

(Received November 13, 1969)

Manuscripts Accepted for Publication in the *Journal of Applied Psychology*

Employee Reactions to Job Characteristics: J. Richard Hackman*, and Edward E. Lawler III: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.

Dimensions of Visual Information Displays: Arthur I. Sigel*, and M. A. Fischl: Applied Psychological Services, Incorporated, Science Center, Wayne, Pennsylvania 19087.

Effect of Individual, Group, and Intergroup Relations on Attitudes Toward a Management Development Program: Clayton P. Alderfer*: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.

Characteristics of Violators of Formal Company Rules: Frans Mulder*: Schubertlaan 6, ENSCHEDE, The Netherlands.

Validity of Importance: Milton R. Blood*: Department of Psychology, University of California at Berkeley, Berkeley, California 94705.

Job Analysis Method for Health Related Professions: A Pilot Study of Physical Therapists: Neil S. Dumas*, and John E. Muthard: National Center for Health Services Research and Development, Department of Health, Education, and Welfare, Rockville, Maryland 20852.

Application Blank: How Honest are the Responses?: Irwin L. Goldstein*: Department of Psychology, University of Maryland, College Park, Maryland 20742.

Relationship Between Task Difficulty and the Criterion: Should We Measure Early or Late?: Edward Levine, and Joseph Weitz*: Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.

Relation of a Trainability Index to T-Group Outcomes: Christopher J. Pino*: Department of Psychology, D'Youville College, 320 Porter Avenue, Buffalo, New York 14201.

Distinguishing Between Engineers Differing on General Job Satisfaction Using Linear Discriminant Function Analysis: Donald A. Wood*: Graduate School of Business, Indiana University, Bloomington, Indiana 47401.

Effect of Personal Signature on the Initial Rate of Return of a Mailed Questionnaire: Mary B. Kawash, and Lawrence M. Aleamoni*: Measurement and Research Division, University of Illinois, 507 East Daniel Street, Champaign, Illinois 61820.

Performance and Knowledge of Results as Determinants of Goal Setting: L. L. Cummings*, Donald P. Schwab, and Marc Rosen: Graduate School of Business, University of Wisconsin, 1155 Observatory Drive, Madison, Wisconsin 53706.

---

* Asterisk indicates author for whom address is supplied.

(Continued on page 150)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 110–113

EFFECTS OF IRRELEVANT STIMULI ON THE PROCESSING OF INFORMATION IN COMPLEX DISPLAYS[1]

IRWIN L. GOLDSTEIN[2]
University of Maryland

AND

JOHN C. ALLEN
University of Tennessee

The present study investigated the effects of irrelevant stimuli on *S*s' ability to organize information by using response categories. Irrelevant stimuli were found to increase response time and errors for all conditions, but more so for certain specific response categories where interference effects could be expected.

The use of response categories refers to the response labels by which observers are asked to group and organize stimuli which appear in complex displays. Several experiments (Long, Henneman, & Garvey, 1960; Long, Reid, & Henneman, 1960; Reid, Henneman, & Long, 1960) produced results that showed increased frequency of correct identifications with increases in category restrictions. An increase in category restrictions is generally accepted to mean an increase in the number of response categories so that they are more descriptive of the specific characteristics of the stimuli. For example, total geometric objects are not as restricted as squares, triangles, circles, and crescents. Generally, improvement in performance was explained by stating that the categories restricted the response alternatives thereby eliminating competing responses and indicating the discriminable aspects of the stimuli. Another study (Goldstein, 1966) found that restricting the response categories helped observers organize stimuli from complex displays when the display contained relatively few objects, but hindered performance when large numbers of objects were presented. The displays used in that study consisted of 35-mm. slides of a constructed model airport with the relevant stimulus objects being cars, trucks, airplanes, and jet planes. However, there were other background objects like radar towers and gas tanks which *S* was instructed to ignore while he grouped relevant stimuli into the proper response categories. The effects of the irrelevant stimuli were not assessed in that study. Yet, it is now apparent (Goldstein, Johnston, & Howell, 1969; Kibler, 1965) that the observer in the man–machine system is going to have to monitor more and more complex displays which in many cases means that he will have to process relevant information intermixed with irrelevant stimuli.

Irrelevant stimuli are often defined as that class of stimuli to which no overt response is required. Unfortunately, the effects of irrelevant stimuli on performance has not been thoroughly investigated and some of the data obtained has been contradictory. Using response time as the dependent variable one investigator (Archer, 1954) did not find any effect as irrelevant stimuli increased while other investigators (Crawford, 1962; Eriksen, 1955; Rabbitt, 1964) have noted significantly longer response times when irrelevant stimuli were added to the display. Studies using errors (identification, counting, etc.) as a dependent variable in the investigation of effects of irrelevant information are quite sparse, though research in the field of concept formation (Bourne & Restle, 1959) should lead one to the view that the presence of irrelevant stimuli probably leads to errors in counting, concept identification, etc.

The present study investigated the effects of irrelevant information as measured by both response time and counting errors. More important, the relationship between irrelevant information and response organization was investigated with the hope of pinpointing some of the conditions and reasons why irrelevant information sometimes has a detrimental effect upon human performance.

---

[1] Computer time was supported in full through the facilities of the Computer Science Center at the University of Maryland. The research was supported by a grant to the Center for Language and Cognition from the Biomedical Science Support Committee of the University of Maryland.

This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Irwin L. Goldstein, Department of Psychology, University of Maryland, College Park, Maryland 20740.

METHOD

Subjects. The *S*s were 40 undergraduate volunteers divided equally into the relevant or the relevant-irrelevant condition. The *S*s received course credit for their participation.

Apparatus. The displays used in this experiment were projections of 35-mm. slides taken of four relevant stimulus classes (squares, triangles, crescents, and circles). The displays were made by randomly placing the objects in a 64-cell matrix. The lines defining the matrix did not appear in the final slides.

The stimulus dimension for the relevant group was varied by changing the number of each of the four relevant stimulus objects appearing in any scene. The three levels of the stimulus complexity dimension were defined as follows: A low level (C_1) included 1, 2, or 3 samples of each relevant stimulus class; an intermediate level (C_2) included 4, 5, or 6 stimuli from each class; a complex level (C_3) included 7, 8, or 9 stimuli from each class. The stimulus dimension for the relevant-irrelevant group was exactly the same for the relevant stimuli. In addition, there were two types of irrelevant objects on the display; pentagons and "nothingons." (A nothingon consisted of two circles which slightly overlap each other.) The number of irrelevant stimuli on the display corresponded to the levels of complexity, that is, for the low level (C_1) 1, 2, or 3 of both the pentagon and the nothingon.

Procedure. The *S*s in the relevant group were given instructions which corresponded to three levels of specificity of response classification. For the reponse category designated as R_1 (one subclass) *S* was required to "tell *E* the total number of geometric objects on the display. These include squares, triangles, circles, and crescents." For the response subclass designated as R_2 (two subclasses), *S* was required to "tell *E* the total number of rounded objects on the displays (these include circles and crescents) and the total number of sided objects (these include triangles and squares)." For the response subclass designated as R_4 (four subclasses), *S* was required to "tell *E* the total number of squares, the total number of triangles, the total number of circles, and the total number of crescents." The *S*s in the relevant-irrelevant group were given the exact same instructions except that they were told there would be two other objects, nothingons, and pentagons, which they were to ignore. The *S*s in both groups were given several practice trials and each *S* was tested to be certain that he could easily discriminate the different stimuli from each other. The *S*s in each group were read the instructions for one of the three response classes and were then shown nine slides, three for each level of stimulus complexity. The same procedure was followed for the second and third response subclasses. The order of the three response subclasses was randomized for each *S* as were the nine slides shown for each of the subclasses. Thus, the relevant versus relevant-irrelevant factor was a between-group variable while the response categories and stimulus complexity were within-group variables.

TABLE 1

MEAN RESPONSE TIME (IN SEC.) FOR THE RELEVANT GROUP AND THE RELEVANT-IRRELEVANT GROUP

Stimulus dimension	Response categories		
	1 Subclass (R_1)	2 Subclasses (R_2)	4 Subclasses (R_4)
Relevant group			
Simple (C_1)	3.96	6.28	6.85
Medium (C_2)	7.80	13.90	16.13
Complex (C_3)	13.72	21.66	24.62
Relevant-Irrelevant group			
Simple (C_1)	6.14	9.63	9.55
Medium (C_2)	13.22	20.62	21.40
Complex (C_3)	20.91	28.37	32.55

RESULTS AND DISCUSSION

Mean response time. Table 1 summarizes the mean response times (in sec.) for the 20 *S*s in the relevant group and the 20 *S*s in the relevant-irrelevant group. An analysis of variance was used to test the significance of the various effects (Winer, 1962, p. 319). The *F* ratios for the stimulus dimension ($F = 499.75$, $df = 2, 76$) and for the response dimension ($F = 161.39$, $df = 2, 76$) were significant at $p < .001$. As can be seen from Table 1, there was an increase in the time scores as both the stimulus complexity and the number of response categories increased. A significant stimulus by response category interaction ($F = 41.27$, $df = 4, 152$, $p < .001$) indicated that the more complex stimuli as compared to the simple stimuli took proportionally longer to classify using the more restrictive response categories (i.e., the four subclass category). These data confirmed earlier studies (Goldstein, 1966) indicating that it obviously required more time to count more objects and more important that the increased discrimination required by the more restrictive response subclasses led to a longer response time. That is, *S*s in the more restrictive response subclass (R_4) counted and named separately all four geometric objects instead of simply counting the total number of geometric objects (R_1). It certainly appears possible to conclude on the basis of the studies completed that regardless of the beneficial effects of response categories in decreasing the number of classification errors, it will take longer processing time for these more complex judgments.

TABLE 2

MEAN ERROR SCORES FOR THE RELEVANT GROUP AND THE RELEVANT-IRRELEVANT GROUP

Stimulus dimension	Response categories		
	1 Subclass (R_1)	2 Subclasses (R_2)	4 Subclasses (R_4)
Relevant group			
Simple (C_1)	.05	.07	.10
Medium (C_2)	.18	.57	.32
Complex (C_3)	1.00	1.47	1.42
Relevant-Irrelevant group			
Simple (C_1)	.10	.23	.15
Medium (C_2)	.57	1.47	.67
Complex (C_3)	1.53	2.25	1.33

A significant relevant-irrelevant effect ($F = 24.89$, $df = 1, 38$, $p < .001$) indicated that it took longer to process the information when irrelevant stimuli were present. These data can be added to a growing body of literature (Crawford, 1962; Eriksen, 1955; Rabbitt, 1964) that supports the general conclusion that the presence of irrelevant stimuli does add to the processing time. In the present study, a significant stimulus complexity by relevancy interaction ($F = 9.73$, $df = 2, 76$, $p < .001$) indicated that as stimulus complexity increased, the relevant-irrelevant group performed proportionally worse than the relevant group. Berlyne (1958) has suggested that the human acts as a communication channel with a limited capacity for receiving increases in information. As the amount of relevant stimuli to be processed increases along with a like increase in the number of irrelevant stimuli, man is going to require proportionally longer and longer time to process the information. It appears that man is not able to select at the peripheral level what information will pass through and what shall be blocked without a severe cost in processing time. The term irrelevant information may be valid only from *E*'s point of view in setting up the experiment. From *S*s' exposure, it is difficult to consider anything having a detrimental affect on his performance as irrelevant. The three-way Relevancy × Response × Stimulus interaction was not significant ($p > .05$).

Mean error scores. Table 2 summarizes the mean error scores for both the relevant and irrelevant group. An error was defined as an underestimation or overestimation of the true number of relevant stimulus objects. For the purpose of the following analyses, these scores were treated on an absolute basis, disregarding sign.

As can be seen from an examination of Table 2, the number of errors increased as stimulus complexity increased ($F = 182.06$, $df = 2, 76$, $p < .001$). In other words, there were more errors with an increased number of objects, a result which should be expected. Also, the relevant-irrelevant effect was significant ($F = 15.50$, $df = 1, 38$, $p < .001$). This main effect generally indicated that *S* made more errors (see Table 2) when irrelevant objects were present. An even more important aspect of this part of these data are the interesting interactions that showed the most detrimental effect occurred for the two-response category in the relevant-irrelevant group. This effect can be observed by examining the significant response ($F = 12.99$, $df = 2, 76$, $p < .001$), Stimulus × Response interaction ($F = 4.40$, $df = 4, 152$, $p < .01$), and Response × Relevancy interaction ($F = 4.04$, $df = 2, 76$, $p < .05$). The three-way Relevancy × Response × Stimulus interaction was not significant ($p > .05$). Table 3 shows that performance became progressively worse for the R_2 response as compared to R_4 response as the stimuli became more and more complex. However, as can be seen by an examination of Table 4, proportionally more of the errors occurred in the relevant-irrelevant group with use of Response Category 2 and 1. Thus, performance was generally worse for those *S*s which were

TABLE 3

MEAN ERROR SCORES AS A FUNCTION OF STIMULUS COMPLEXITY AND RESPONSE CATEGORIES

Stimulus dimension	Response categories		
	1 Subclass (R_1)	2 Subclasses (R_2)	4 Subclasses (R_4)
Simple (C_1)	.08	.15	.13
Medium (C_2)	.38	1.02	.49
Complex (C_3)	1.27	1.86	1.38

presented with irrelevant information and that decrement in performance was proportionally worse for *S*s in the relevant-irrelevant group when they used Response Category 2 and 1.

This generally higher error rate for groups which must process information in the presence of irrelevant stimuli shows that for this task *S*s longer response times for the irrelevant conditions did not compensate *S*s and lead to performance similar to relevant *S*s. Thus, irrelevant stimuli add to processing time and also contribute to a higher error rate. An examination of the R_2 and R_1 category for the relevant-irrelevant group suggests one way that irrelevant stimuli may add to the error rate. The irrelevant objects were sided objects (pentagon) and round objects (nothingon) and it appears quite likely that these two irrelevant objects interfered with the two-response category called sided versus round. It is also interesting to speculate that these two irrelevant objects caused some difficulty with the one-response category called total geometric objects since these objects were indeed geometric objects. However, in the four-response category group, each of the four relevant objects are identified by their specific name (e.g., triangle, crescent). The two irrelevant objects are identified with different names and shapes from the relevant objects and might not be expected to cause as much interference. Indeed, an examination of Table 4 indicates that the error scores for the relevant group and the relevant-irrelevant group were quite similar for the four-response category. Certainly, there is now a possibility that our original belief that response categories can be used to help organize information must now be reconsidered to the extent to which there is irrelevant information present which may interfere with the use of the categories.

In any case, the irrelevant stimuli generally retarded performance. As displays tend to become more complex with the operator expected to pick out the pertinent (relevant) information from the background (irrelevant) stimuli, the data presented in this study indicate it will be necessary to consider ways in which to overcome the presence of "irrelevant stimulation."

TABLE 4

MEAN ERROR SCORES AS A FUNCTION OF RESPONSE CATEGORIES AND RELEVANCY

Relevancy conditions	Response categories		
	1 Subclass (R_1)	2 Subclasses (R_2)	4 Subclasses (R_4)
Relevant	.41	.70	.61
Relevant-Irrelevant	.73	1.32	.72

REFERENCES

ARCHER, J. E. Identification of visual patterns as a function of information load. *Journal of Experimental Psychology*, 1954, **48**, 313–317.

BERLYNE, D. E. The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology*, 1958, **55**, 289–296.

BOURNE, L. E., & RESTLE, F. Mathematical theory of concept identification. *Psychological Review*, 1959, **66**, 278–296.

CRAWFORD, A. The perception of light signals: The effect of the number of irrelevant lights. *Ergonomics*, 1962, **5**, 417–428.

ERIKSEN, C. W. Partitioning and saturation of visual display and efficiency of visual search. *Journal of Applied Psychology*, 1955, **39**, 73–77.

GOLDSTEIN, I. L. The effects of stimulus complexity and restrictive response subclasses on observer responses. *Journal of Experimental Psychology*, 1966, **71**, 104–108.

GOLDSTEIN, I. L., JOHNSTON, W., & HOWELL, W. Complex vigilance: Relevant and irrelevant signals. *Journal of Applied Psychology*, 1969, **53**, 45–48.

KIBLER, A. W. The relevance of vigilance research to aerospace monitoring tasks. *Human Factors*, 1965, **7**, 93–99.

LONG, E. R., HENNEMAN, R. H., & GARVEY, W. D. An experimental analysis of set: The role of sense modality. *American Journal of Psychology*, 1960, **73**, 563–567.

LONG, E. R., REID, L. S., & HENNEMAN, R. H. An experimental analysis of set: Variables influencing the identification of ambiguous visual stimulus-objects. *American Journal of Psychology*, 1960, **73**, 553–562.

RABBITT, P. M. A. Ignoring irrelevant information. *British Journal of Psychology*, 1964, **55**, 403–414.

REID, L. S., HENNEMAN, R. S., & LONG, E. R. An experimental analysis of set: The effects of categorical restriction. *American Journal of Psychology*, 1960, **73**, 568–572.

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received November 14, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 114–118

INTERACTION OF ENVIRONMENT AND PERSONALITY:

A TEST OF HOLLAND'S THEORY[1]

CHARLES F. ELTON[2]

University of Kentucky

Holland's assumption that personality variables interact with environment was tested by the following hypotheses: (*a*) For males who leave engineering ($n = 27$) there is a personality change different from that found among those who remain in engineering ($n = 25$). (*b*) For males who leave or remain in engineering, the personality development is different from that of males in a 2-yr. college ($n = 22$). Two dimensions represented the differences between groups and these were labeled Intellectual Independence and Practical Inclination. Transfers to arts and sciences from engineering were described as becoming more realistic, nonjudgmental, intellectual, liberal, and skeptical of orthodox religious beliefs. Two-year students showed greater variability in their personality change scores and a loss in their liking for reflective and abstract thought.

Holland (1966) proposed the possibility of characterizing persons according to one of six personality types: realistic, intellectual, social, conventional, enterprising, and artistic. The more closely a person resembles a particular personality type, the more likely it is that he will exhibit personal traits and behaviors associated with that type. Furthermore, the environment in which a person lives may be described according to six model environments which are analogous to the personality types. An essential aspect of Holland's theory is that outcomes can be predicted on the basis of knowledge concerning the congruence between personality type and environment. For example, a person characterized as an enterprising personality type would be expected to be most productive while working in an enterprising environment. Another prediction which emerges from the interaction of personality and environment is related to vocational stability. Students whose personality characteristics are incongruent with their vocational choice would be predicted to change majors. Data reported by Elton and Rose (1966) support this prediction.

A more crucial test of the theory that personality type interacts with environment is found in the hypothesis that a student who changes his major should change in his personality development; while one who remains in the same major will not change in the same way. A corollary hypothesis is that an environment allowing congruence between personality and vocational choice will produce personality changes which are different from those produced by an environment prohibiting congruence between personality and vocational choice.

The purpose of this study is to investigate the following two hypotheses:

1. For males who leave engineering there is personality change different from that found among those who remain in engineering.
2. For university males who remain in engineering, as well as those who change majors, the personality development is different from that of males in a 2-yr. college.

METHOD

The Omnibus Personality Inventory (OPI), Form C was administered to the 1965 freshman classes at the University of Kentucky and at the Ashland Community College in Ashland, Kentucky. From lists of students who were continuously enrolled four semesters, excluding the summer term, three samples of 50 males were randomly selected. Two samples represented University of Kentucky students, and one was from the Ashland Community College. One of the University of Kentucky samples consisted of students who transferred to the College of Arts and Sciences after completing two semesters in the College of Engineering. The second was composed of students who completed four semesters in the College of Engineering. The Ashland Community College students were those enrolled in the 4-yr. college transfer program. The students in each sample were invited by letter to take the OPI

---

[1] This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to the author, College of Education, University of Kentucky, Lexington, Kentucky 40506.

during the last month of their fourth semester in college. As an incentive to participate in this testing, each student was offered $2 for his cooperation. Data were finally collected for 27 arts and sciences transfers, 25 engineering majors, and 22 Ashland males.

Twelve OPI scale scores of the total University of Kentucky freshman population who enrolled in 1965 and 1966 were subjected, by sex, to a principal components analysis. Four scales, Developmental Status, Social Maturity, Repression-Suppression, and Nonauthoritarianism were omitted from analysis because of their large item overlap with the other scales. A complete description of the OPI scales and their development may be found in the OPI research manual (Center for the Study of Higher Education, 1962).

Because of the relationship between size of N and stability of factor scores, the factor weights derived from the 1965 and 1966 University of Kentucky factor analysis were used to estimate factor scores for the students in each of the three samples in this study. This factor structure and the OPI scale weights are reported in Rose and Elton (1968). To estimate factor scores with a mean of 50 and a standard deviation of 10, the procedure suggested by Lawley and Maxwell (1963) was followed.

Alteration in the factor structure between the first and second administration of the OPI could shift the meaning of the factor scores. To show stability in factor structure between freshman and senior OPI scale scores, Stewart (1964) described an innovative use of canonical correlation. Because, in this study, factor scores rather than scale scores were employed, it was possible to eliminate the possibility of an altered meaning in factor score by applying identical factor weights to the 1965 and 1967 OPI scale scores for each student.

The 1965 OPI factor scores were tested for the significance of mean differences between groups by a t test and none of these values were significant at the .05 level of confidence.

For each student in this study, five estimated factor scores were available from the 1965 administration of the OPI, and an equal number of factor scores were available from the 1967 administration of the OPI. It seemed desirable to develop a measure of change which would reflect both the amount and direction of change in each student's factor scores between these two time periods. Tucker, Damarin, and Messick (1966) propose one method for obtaining a base-free measure of change and this method was used. This measure of change insures that observed change cannot be attributed to regression effects. Consequently, the five 1965 OPI factor scores and the five 1967 OPI factor scores were reduced to five OPI factor change scores for each student. These scores were obtained by subtracting the freshman score from the sophomore score.

Discriminant analysis was used to answer the question: "Are the factor change scores between these three samples of males different?"

RESULTS

TABLE 1

MEAN OPI FACTOR SCORES AND DISCRIMINANT WEIGHTS

OPI factor	Arts and sciences[a]	Ashland[b]	Engineering[c]	Discriminant weight	
				λ1	λ2
Nonconformity	17.77	12.18	17.96	.05	−.02
Scholarly Orientation	9.07	−4.43	6.29	.10	−.10
Nonauthoritarianism	13.01	20.55	9.30	−.08	−.10
Social Discomfort	12.59	10.80	11.16	−.02	−.04
Masculine Role	17.78	13.06	21.23	.02	.06

Note.—OPI = Omnibus Personality Inventory.

[a] $n = 27$.

[b] $n = 22$.

[c] $n = 25$.

Table 1 presents the mean factor change scores between the three samples of males. The discriminating power of the predictors was tested by chi-square with $p(k - 1)$ degrees of freedom, where p is the number of variables and k is the number of groups. This chi-square value was 61.57 which, with 10 degrees of freedom, was significant at the .001 level. Each discriminant dimension was tested by chi-square to evaluate its separate contribution to discrimination. On the first dimension, which accounted for 88% of the variation between groups, the chi-square value was 56.76, significant at the .001 level with 6 degrees of freedom $(p + k - 2)$. The second discriminant dimension, accounting for 22% of the variation between the groups, produced a chi-square value of 10.39, which, with 4 degrees of freedom $(p + k - 4)$, was significant at the .05 level of confidence. Additional dimensions failed to achieve significance at the .05 level by the chi-square test.

The F statistic was used to test the equality between group means of the OPI factor change scores. Students transferring to the College of Arts and Sciences and those remaining in engineering obtained significantly higher change scores on the variable of Scholarly Orientation $(p < .01)$ than did students at the Ashland Community College. The arts and sciences transfers, compared to those remaining in engineering, obtained a significantly higher score on the factor of Nonauthoritarianism $(p < .05)$. Other group mean change scores did not differ significantly at the .05 level of confidence.

The effectiveness of factor change scores in classifying the students into three groups is shown in Table 2. The best discrimination was obtained for the Ashland students, the poorest

TABLE 2

ACTUAL AND PREDICTED GROUP MEMBERSHIP

Group	Arts and sciences	Ashland	Engineering	% hits
Arts and sciences	14	4	9	52
Ashland	0	20	2	90
Engineering	6	2	17	68

for the arts and sciences transfers, with the engineers falling between these two extremes.

Figure 1 is a plot of the mean discriminant function values on the two significant dimensions.

The horizontal dimension in Figure 1 represents in graphic form the major differences found between the three student samples in personality change scores. The contributors to differentiation on the first dimension, which is labeled Intellectual Independence, are Scholarly Orientation scores, positively weighted and Nonauthoritarianism scores, negatively weighted. High scores on positively weighted predictors and low scores on negatively weighted predictors produce a high dimension score; a low score on a dimension is the result of low scores on positively weighted predictors and high scores on negatively weighted predictors. The discriminant weights and mean change scores given in Table 1 indicate that Ashland students should have a low score on Intellectual Independence and students at the University of Kentucky should have a high dimension score. The mean dimension scores plotted in Figure 1 support this prediction.

The major contributors to the second dimension labeled Practical Inclination in Figure 1 are Scholarly Orientation and Nonauthoritarianism scores, both negatively weighted, and Masculine Role, positively weighted. The implication of this separation is that students transferring to arts and sciences from engineering change differently compared to those remaining in engineering.

DISCUSSION

The finding of differential change in the Nonauthoritarianism score for *S*s transferring to arts and sciences compared to those remaining in engineering lends support to Holland's assumption that personality type interacts with environment. In contrast to *S*s in engineering, the transfers to arts and sciences may be described as becoming more realistic, nonjudgmental, intellectually liberal, and skeptical of orthodox religious beliefs. Noteworthy is the fact that significant differences were not found in the amount of change in Nonconformity, Scholarly Orientation, Social Discomfort, and Masculine Role scores.

The environmental differences between students remaining in engineering and those transferring to arts and sciences are presumed to be those associated with faculty, curricular, and peer influences. Not collected were data which would allow inferences about those aspects of the environment related to observed change in personality factor scores.

Two-year students differed from university students in their personality change scores. This differences has both quantitative and qualitative aspects. The quantitative difference may be observed in the variability between groups in mean change scores. The range in change scores for the Ashland *S*s was 24.98; for engineers it was 14.94; for the arts and sciences transfers it was 8.71. The Ashland *S*s had almost three times the variability of the transfers to arts and sciences and more than one and a half times the variability of engineers.

The pattern of score change illustrates the qualitative differences. On one variable—Scholarly Orientation—2-yr. *S*s earned the lowest mean change score, in fact, a negative score. On another variable—Nonauthoritarianism—2-yr. students earned the highest mean change score. Nonauthoritarianism did not add significantly to the discrimination between 2- and 4-yr. students because of the variance accounted for by Scholarly Orientation. It is evident, though, that the mean Ashland change score on Nonauthoritarianism has a meaningful effect, but opposite to that produced by its Scholarly Orientation score.

The differences found among 2-yr. college students compared to those in the other two groups are in the direction predicted by Holland's theory. It is presumed that the environmental differences between the Ashland campus and the University were instrumental

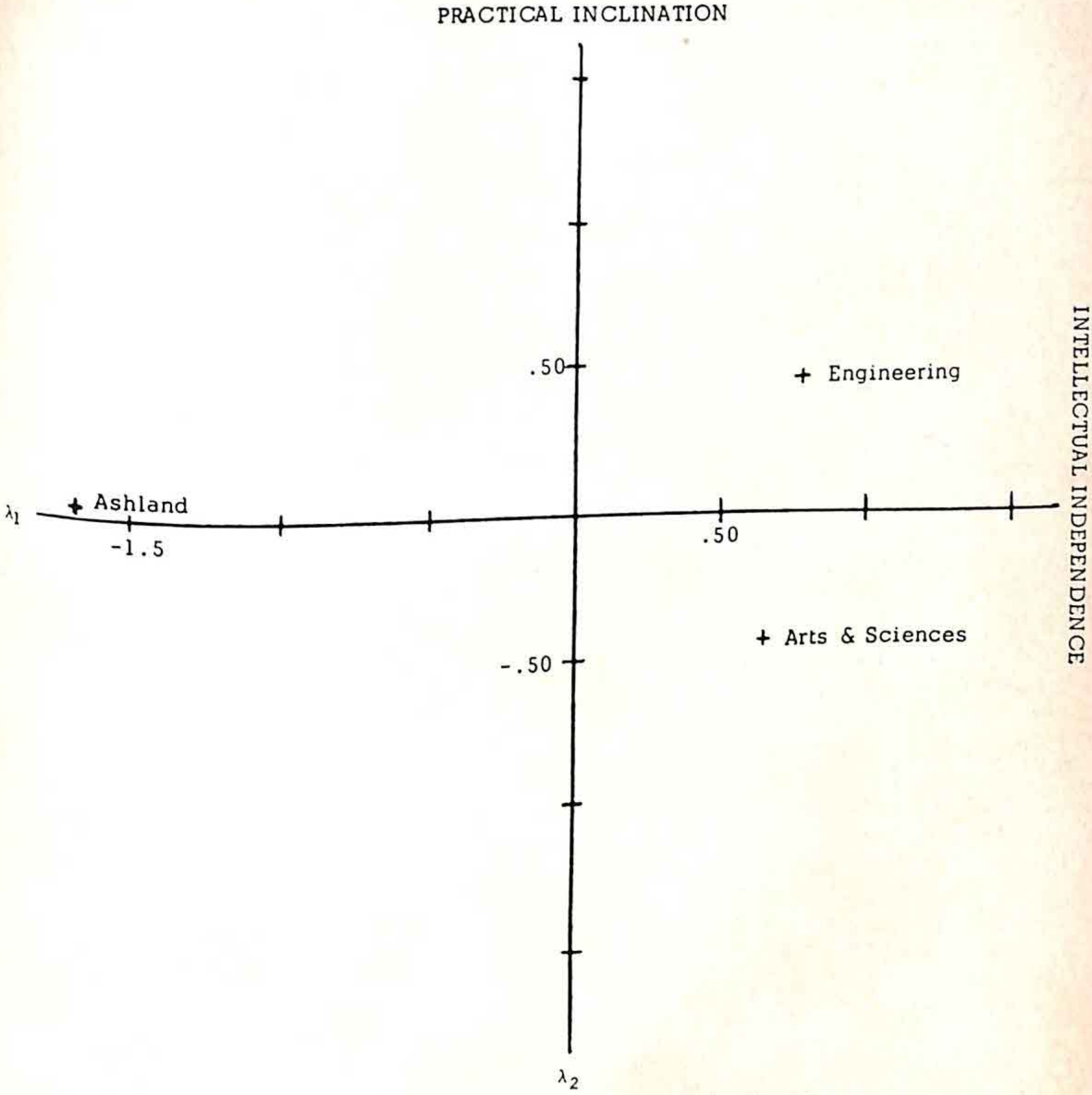


Fig. 1. Mean values of dimensions in discriminant space.

in producing the observed differences. More specifically, Ashland males did not have an opportunity to contrast and compare their individual personality characteristics with curricular requirements, peers, or faculty who might serve as vocational role models. In the absence of such models, it is not surprising that Ashland *S*s showed the largest variability in personality development. Or, stated another way, personality change among students enrolled in a 2-yr. college would be expected to be influenced less by faculty, peer, and curricular influences than are changes found among students enrolled in a traditional college program.

One aspect of the difference found between Ashland and University *S*s invites a speculative comment. This difference is portrayed by the dimension of Intellectual Independence in Figure 1. Could it be that the transfer program at a 2-yr. college produces a change in personality characteristics which is antithetical to curricular objectives which stress intellectual independence? If so, it would be predicted that Ashland students with negative Scholarly Orientation change scores who transferred to a 4-yr. science program would fail in larger numbers than would those with positive change scores. The significance of this speculation is enhanced by data reported by Furneaux

(1962) which indicate that personality characteristics operate as moderating variables in academic achievement.

Plant and Minium (1967) showed that change in personality scores was related to ability. The mean American College Test (ACT) score for the Ashland *S*s was 20; it was 24 for the transfers to arts and sciences and those remaining in engineering. The correlation between Scholarly Orientation and ACT was, however, only .08. Therefore, it does not seem probable that ability differences between 2- and 4-yr. students account for the observed differences in Scholarly Orientation scores. Additional evidence which supports the independence of ability and personality change scores was found in the significance difference in Nonauthoritarianism scores between arts and sciences transfers and engineers. The mean ACT scores were identical for these two groups of students.

Consistent with the nature of many exploratory studies the present investigation suggests a number of tentative conclusions. The findings cannot be presumed to be universal; the small number of *S*s and the absence of cross-validated data mitigate against overgeneralization. The data do emphasize the heuristic value of relating observed change in personality test scores to Holland's theory of vocational development.

REFERENCES

CENTER FOR THE STUDY OF HIGHER EDUCATION. *Omnibus Personality Inventory research manual.* Berkeley: Author, 1962.

ELTON, C. F., and ROSE, H. A. Within-university transfer: Its relation to personality characteristics. *Journal of Applied Psychology*, 1966, **50**, 539–543.

FURNEAUX, W. D. The psychologist and the university. *Universities Quarterly*, 1962, **17**, 33–47.

HOLLAND, J. L. *The psychology of vocational choice.* Waltham: Blaisdell, 1966.

LAWLEY, D. N., and MAXWELL, A. E. *Factor analysis as a statistical method.* London: Butterworths, 1963.

PLANT, W. T., and MINIUM, E. W. Differential personality development in young adults of markedly different aptitude levels. *Journal of Educational Psychology*, 1967, **58**, 141–152.

ROSE, H. A., and ELTON, C. F. Acceptors and rejectors of counseling. *Journal of Counseling Psychology*, 1968, **15**, 578–580.

STEWART, L. H. Change in personality test scores during college. *Journal of Counseling Psychology*, 1964, **11**, 211–220.

TUCKER, L. R., DAMARIN, F., and MESSICK, S. A base-free measure of change. *Psychometrika*, 1966, **31**, 456–473.

(Received November 17, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 119–122

JOB OFFER ACCEPTANCE BEHAVIOR AND REINFORCEMENT[1]

JOHN M. IVANCEVICH AND JAMES H. DONNELLY[2]

University of Kentucky

The effects of a positive reinforcement program on a hypothesized indicator of dissonance (college seniors rescinding a job offer acceptance after previously accepting) was investigated in two separate companies. A total of 196 *S*s (experimental group) were provided reinforcement from those firms they decided to join, while another 196 *S*s (control group) received no reinforcement. Those *S*s receiving reinforcement rescinded job offers at a significantly lower rate than those *S*s not receiving reinforcement.

During the last decade the interest, debate, and controversy over such concepts as balance, congruity, and dissonance has not subsided. If anything, the interest in these concepts first displayed by psychologists has generated more interest recently among researchers interested in communications, consumer behavior, and organizational analysis.

The focus of the attitude change theories of Osgood, Suci, and Tannenbaum (1957); Abelson and Rosenberg (1958); Katz and Stotland (1959); and Rosenberg (1960) involved the principle of cognitive consistency. The approach to interpersonal behavior suggested by Heider (1946), as well as the approaches of Newcomb (1953), and Festinger (1957, 1964) also emphasize that individuals tend toward maintaining cognitive balance, symmetry, or consonance.

Numerous research efforts (Brehm & Cohen, 1959; Cohen, Brehm, & Latané, 1959; Ehrlich, Guttman, Schönback, & Mills, 1957; Engel, 1963; Freedman, 1965; Fromkin, 1968; and Oshikawa, 1968) have attempted to relate post-decision behavior to cognitive dissonance. The main theme of most of these empirical efforts is that after a person makes a decision, such as purchasing a home or an automobile, there is often a lack of consistency or harmony among the person's cognitions.

Post-decision doubts are likely to occur when (*a*) the decision is important to the individual psychologically and/or financially (Brehm, 1960; Culbertson, 1957), (*b*) a number of desirable alternatives are available (Anderson, Taylor, & Holloway, 1966), and (*c*) available alternatives are qualitatively dissimilar (Brehm & Cohen, 1959).

The present research study attempts to relate the effects of positive reinforcement on a hypothesized indicator of post-decision dissonance. Specifically, it reports the findings of a recent investigation concerning the tendency of some college seniors to back out (hypothesized indicator of dissonance) of a job offer acceptance after formally notifying a company that they will begin their careers with the firm immediately after graduation. The first job decision of the college graduate fully meets the post-decision dissonance criteria discussed by psychologists. First, it is an important psychological and financial decision for the graduating senior. Second, most college seniors receive a number of job offers from competing organizations. Finally, the job offers are generally from organizations in different industries, from different locations, and involve different responsibilities.

The main concern of the researchers in the present study is not to resolve the heated debate surrounding cognitive dissonance (Chapanis & Chapanis, 1964; Janis & Gilmore, 1965), but rather to determine if a program of positive reinforcement reduces back-out frequencies (assumed indicator of dissonance) in organizations. Consequently, the operational proposition investigated is:

> The number of college graduates backing out after formally accepting a job offer (e.g., through telephone conversation and/or letter) will be reduced if positive post-decision reinforcement is provided by the hiring company.

---

[1] This article was accepted for publication by the outgoing editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to James H. Donnelly, Department of Business Administration, University of Kentucky, Lexington, Kentucky 40506.

TABLE 1

AVERAGE BACK-OUT PERCENTAGES

No. of firms[a]	Average % of back outs
7	<1
5	1–2.99
2	3–4.99
1	5–7.99
2	8–11.99
1	12–15.99
1	>16

[a] Each of the 19 companies hired more than 125 college graduates in 1967. The range of college graduates hired was from 125 to 1,650. These figures include undergraduate and graduate degrees.

METHOD

Procedure

Nineteen large industrial corporations were contacted personally and by questionnaire in 1966 and 1967 to determine the number of college graduates who, after formally accepting a position, later rejected the position or never showed up on the job. This inquiry indicated that, of the 19 participating firms contacted, each had reported some backing out among college graduates. Table 1 indicates the approximate percentage[3] of back outs in the 19 large industrial companies.

This initial inquiry indicated that some companies have a critical back-out problem. For example, one of the companies contacted had a back-out rate which was in excess of 16%. Thus, the researchers assumed that back-out rates could be reduced significantly in those companies with a back-out rate which is greater than 5% if positive reinforcement on the part of the firm was initiated during the postdecision period.

Participating Companies

Two companies from the 19 originally studied that had back-out rates of approximately 10% agreed to participate in this study. Company A is located in New York and during 1967 had hired through the main corporate offices 131 college seniors for nontechnical positions (e.g., business administration and liberal arts degree holders). Company B which is located in the Chicago area had, during 1967, hired 140 technical college graduates (e.g., engineering and physical science holders). In the initial inquiry (i.e., when the 19 companies were contacted), it was determined that Company A's back-out rate was 10.4% and Company B's back-out rate was 10.9%.

Experimental and Control Groups

During 1968 and through June of 1969, Company A received 186 formal job offer acceptances from undergraduates and, during this same period, Company B received 206 formal job offer acceptances. One of every two acceptances at Company A was classified into a control group and the other into an experimental group. For example, suppose that on April 2, College Seniors 1, 2, and 3 (CS_1, CS_2, and CS_3) formally accepted. Thus, CS_1 was placed in the control group, CS_2 into the experimental group, and CS_3 into the control group. Then on April 6, CS_4 and CS_5 formally accepted. In this case, CS_4 was placed into the experimental group and CS_5 into the control group. This same classification procedure was followed for all of the 186 formal job acceptances at Company A and the 206 acceptances at Company B.

The researchers were interested in whether those accepting offers had other possible job alternatives. A 20% random sample of those accepting and beginning their careers with Company A or Company B and of those accepting and later backing out indicated that each of the *S*s had at least one other job offer. The average number of offers for those accepting and reporting to work was 3.6, while the average number of offers of those accepting and backing out was 3.3.

Members of the control group at each company received some company literature and confirmation that their notice of acceptance had been received by the personnel department. The experimental groups, however, received three randomly spaced personal telephone calls from the personnel department which stressed positive factors about the company and the astuteness of the college senior's company choice. The first of these randomly spaced calls was made immediately after the formal acceptance was received in the personnel department. The personnel department participants (i.e., six at Company A and seven at Company B) were instructed by the researchers to emphasize in their telephone conversations such factors as the opportunities for advancement in the firm, the challenge of working in the organization, the positive aspects of the community in which the firm is located, and other similar attributes associated with working for the company. In addition to these randomly spaced telephone calls, the same literature and confirmation sent to the control group members was also sent to the experimental group *S*s.

RESULTS

Table 2 presents the frequencies of backing out among college seniors receiving positive reinforcement (experimental group) and college seniors not receiving positive reinforcement (control group) for Company A. Chi-square analysis was used to test for statistically significant differences.

In Table 3, the frequencies of backing out for the experimental and control groups at Company B are shown. Once again, chi-square analysis was used to determine if a significant difference exists between groups.

---

[3] The figure is stated as an approximation since six firms did not keep accurate records of the number of college graduates backing out of an original job offer acceptance. Originally, 28 companies were contacted by mail, but only 19 expressed an interest to pursue discussions with the researchers. Six of the nine companies who did not want to discuss this topic indicated that the back-out rates in their firms were significantly less than 1% annually and they were not concerned with the backing-out problem.

Examination of Tables 2 and 3 indicates that the positive reinforcement programs utilized in both companies had a significant impact on reducing the back-out rate. The number of college seniors backing out in the experimental groups was significantly less than those control group *S*s who received no postdecision reinforcement. Thus, the empirical data presented in Tables 2 and 3 is consistent with the original proposition.

DISCUSSION

Theoretically, through a carefully planned recruitment program, personnel selectors arrive at a decision about which of the numerous job candidates clearing all of the selection "hurdles" are to receive employment offers. The most qualified candidates supposedly receive these offers. It has been found that some college seniors after formally accepting an offer, will later rescind the previous acceptance. This back-out tendency was reduced in the present study when systematic postdecision reinforcement programs were utilized.

In the control groups in both companies, the back-out percentages during the investigation remained at approximately the same level they had been prior to the study. The experimental groups, however, showed a significant reduction in the tendency to back out. The reinforcement program was equally effective in influencing both technical and nontechnical undergraduates.

Although significant reductions in back-out rates were brought about through the reinforcement strategy utilized, a number of important questions still remain unanswered. First, it was not determined how many telephone calls or what type of reinforcement (i.e., letters plus telephone calls, letters alone, telephone calls alone, or some other similar method) is most effective in reducing the back-out rate. For those organizations who believe they have an excessive back-out rate, it would seem worthwhile to investigate the effects of various reinforcement strategies upon backing out. Also, it was not determined what the best source of reinforcement would be. In this study, the personnel department initiated the reinforcement. Perhaps, the prospective supervisor of the new employee (e.g., the production manager contacts incoming production personnel) would serve as a more effective initiator of reinforcement.

TABLE 2

BACK-OUT FREQUENCIES IN COMPANY A

Group[a]	Did not back out	Did back out	% backing out
Experimental	91	2	2.15
Control	83	10	10.75

Note.—$\chi^2 = 5.68$, $df = 1$, $p < .02$.
[a] $n = 93$ in both groups.

TABLE 3

BACK-OUT FREQUENCIES IN COMPANY B

Group[a]	Did not back out	Did back out	% backing out
Experimental	100	3	2.91
Control	92	11	10.68

Note.—$\chi^2 = 4.90$, $df = 1$, $p < .05$.
[a] $n = 103$ in both groups.

Second, it would be worthwhile to determine whether the positive reinforcement is related in any degree to the first year performance on and/or attitudes about the job of college graduates. This can be accomplished by studying these variables as related to the control group and experimental group *S*s during the first year of employment.

Third, the exact reasons for backing out of a formal job offer acceptance would provide valuable information to personnel departments. The authors of the present study have inferred that backing out is caused by an intense state of dissonance caused by postdecision anxiety. A more precise analysis, however, may find another explanation for the back-out phenomena. Interviews with seniors who have backed out could lead to programs which could, in some firms, almost completely eliminate backing out.

REFERENCES

ABELSON, R. P., & ROSENBERG, M. J. Symbolic psycho-logic: A model of attitudinal cognition. *Behavioral Science*, 1958, **3**, 1–13.

ANDERSON, L. K., TAYLOR, J. R., & HOLLOWAY, R. J. The consumer and his alternatives: An experimental approach. *Journal of Marketing Research*, 1966, **3**, 62–67.

BREHM, J. W. Attitudinal consequences of commitment to unpleasant behavior. *Journal of Abnormal and Social Psychology*, 1960, **60**, 379–383.

BREHM, J. W., & COHEN, A. R. Re-evaluation of choice alternatives as a function of their number and qualitative similarity. *Journal of Abnormal and Social Psychology*, 1959, **58**, 373–378.

CHAPANIS, N. P., & CHAPANIS, A. Cognitive dissonance: Five years later. *Psychological Bulletin*, 1964, **61**, 1–22.

COHEN, A. R., BREHM, J. W., & LATANÉ, B. Choice of strategy and voluntary exposure to information under public and private conditions. *Journal of Personality*, 1959, **27**, 63–73.

CULBERTSON, F. M. Attitude change through role playing. *Journal of Abnormal and Social Psychology*, 1957, **54**, 230–233.

EHRLICH, D., GUTTMAN, I., SCHÖNBACH, P., & MILLS, J. Post-decision exposure to relevant information. *Journal of Abnormal and Social Psychology*, 1957, **54**, 98–102.

ENGEL, J. F. Are automobile purchasers dissonant consumers? *Journal of Marketing*, 1963, **27**, 55–58.

FESTINGER, L. A. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson & Co., 1957.

FESTINGER, L. *Conflict, decision, and dissonance.* Stanford: Stanford University Press, 1964.

FREEDMAN, J. L. Confidence, utility, and selective exposure: A partial replication. *Journal of Personality and Social Psychology*, 1965, **2**, 778–780.

FROMKIN, H. L. Reinforcement and effort expenditure: Predictions of reinforcement theory versus predictions of dissonance theory. *Journal of Personality and Social Psychology*, 1968, **9**, 347–352.

HEIDER, F. Attitudes and cognitive organization. *Journal of Psychology*, 1946, **21**, 107–112.

JANIS, I. L., & GILMORE, J. B. The influence of incentive conditions on the success of role playing in modifying attitudes. *Journal of Personality and Social Psychology*, 1965, **1**, 17–27.

KATZ, D., & STOTLAND, E. A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology of a science.* New York: McGraw-Hill, 1959.

NEWCOMB, T. M. An approach to the study of communicative acts. *Psychological Review*, 1953, **60**, 393–404.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. The measurement of meaning. Urbana, Ill.: University of Illinois Press, 1957.

OSHIKAWA, S. The theory of cognitive dissonance and experimental research. *Journal of Marketing Research*, 1968, **5**, 429–430.

ROSENBERG, M. J. Cognitive reorganization in response to the hypnotic reversal of attitudinal affect. *Journal of Personality*, 1960, **28**, 39–63.

(Received November 19, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 123–125

DISPLAY CHARACTERISTICS IN INSPECTION TASKS[1]

ROBERT C. WILLIGES[2] AND HARRISON STREETER

University of Illinois

Eight inspectors were required to detect pin-hole defects in a visual scanning task. Display type (static vs. dynamic) and display arrangement (random vs. ordered) were combined factorially in a within-*S* design. Static (inspector-paced) displays yielded more defect detections and fewer false alarm errors; whereas, display arrangement did not effect detections or false alarm errors. It was concluded that inspector-paced displays seem to be more beneficial than externally paced displays when rapid scanning rates are required.

Quality control inspection is a task that man is increasingly called upon to perform in modern industrial situations. Inspection is essentially a decision as to whether or not a particular item meets a specified standard. Although this task is common in most industrial manufacturing operations, surprisingly little research has been conducted on factors which affect inspector performance. Harris (1969) provided a taxonomy of the various types of inspection tasks and the common behavorial elements involved in these tasks. It is important to understand factors considered in this taxonomy which affect performance before the capabilities of the human inspector can be adequately established and inspector accuracy increased.

Only recently, such variables as observing responses (Badalamente & Ayoub, 1969), task complexity (Harris, 1966), costs and payoffs (Wallack & Adams, 1969), and defect rate (Harris, 1968) have been submitted to experimental investigation. Characteristics of the display of inspection items and their subsequent effects on inspector performance have not received experimental treatment.

This study dealt with simple visual scanning inspection in which the inspector was required to detect pin-hole defects. The primary purpose of the study was to investigate the effects of display arrangements which could structure the inspection task either by providing a pattern of scanning or by pacing the rate of inspection. It was hypothesized that a visual inspection task which presents an ordered array of items in a dynamic display should provide better inspector accuracy than those in which items are presented randomly or in static mode.

METHOD

Task. The inspection task involved detecting pin-hole defects in ¼-in. diameter transparent yellow plastic discs. The discs were displayed in groups of 600. Seventy-five discs with pin-hole defects appeared randomly within each group.

Each group of discs was displayed in two ways—a static display with inspection discs mounted on a flat 6 × 40 in. light table, and a dynamic display with inspection discs mounted on a moving translucent belt on a drug capsule sorting machine. The belt moved at a constant speed displaying approximately 100 discs at a time and presented the entire group of 600 discs in 20 sec. Both types of displays were lighted by fluorescent tubes mounted underneath the white lucite display surfaces.

Two arrangements of the 600 discs mounted on the displays were employed. One arrangement was completely random with the restriction that no two discs overlap. The other arrangement was ordered in an 8 × 75 matrix on ½-in. centers.

Subjects and design. Eight female undergraduate students at the University of Illinois participated in the experiment as part of a psychology course requirement. None of the *S*s had previous experience as an industrial inspector.

A within-*S* factorial design consisting of two factors, display arrangement and display type, was employed. The two levels of display arrangement were random and ordered configurations. Display type was either a static (inspector-paced) or dynamic (*E*-paced) display of discs.

Procedure. Prior to beginning the inspection task, each *S* was shown examples of disc defects and was instructed that she would be allowed 20 sec. to inspect 600 discs for similar defects. Each *S* then inspected discs in each of the four factorial combinations of display type and display arrangement. Order of presentation of the conditions was counterbalanced across *S*s.

Each *S* was given two inspection trials on each of the four experimental conditions, the first trial serving as practice. To facilitate scoring, a layer of clear acetate was laid over each display and the *S* used a china marker pencil to mark on the acetate over any disc she thought

---

[1] This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Robert C. Williges, University of Illinois, Department of Psychology, Champaign, Illinois 61820.

TABLE 1

AVERAGE FREQUENCY OF DEFECT DETECTIONS AND FALSE ALARM ERRORS OVER BLOCK OF 25 DEFECTS FOR TWO TYPES OF DISPLAYS

Display type	Detection	False alarms
Dynamic	11.08	.88
Static	14.97	.19

was defective. Each *S* completed all of her trials within a single experimental period.

RESULTS

Analyses were conducted on both the number of defects correctly detected and the number of false alarm errors (marking a disc defective when it had no pin-hole defect). Since it was noted that most *S*s did not inspect all the discs under static conditions, the results were analyzed over three blocks of discs, each containing 25 defects. A $2 \times 2 \times 2 \times 3$ repeated-measures analysis of variance design consisting of display type, display arrangement, trials, and blocks was used on both dependent variables.

False Alarms

The only significant effect on the analysis false alarm errors was due to the main effect of display type ($F = 19.50$, $df = 1, 7$, $p < .01$). The average frequency of false alarms are shown in the right portion of Table 1. Overall error rate was low, but more false alarms occurred under dynamic displays then under static displays.

Detections

Significant main effects on detection performance were found due to display type ($F = 16.15$, $df = 1, 7$, $\mathrm{p} < .01$), blocks ($F = 21.54$, $df = 2, 14$, $p < .001$), and trials ($F = 31.03$, $df = 1, 7$, $p < .001$). The Display Type × Blocks interaction ($F = 15.90$, $df = 2, 14$, $p < .001$); the Display Type × Display Arrangement × Blocks interaction ($F = 5.55$, $df = 2, 14$, $p < .05$); and the Display Type × Display Arrangement × Trials × Thirds interaction ($F = 13.26$, $df = 2, 14$, $p < .001$) were also significant.

Figure 1 depicts these significant effects. Static displays resulted in higher frequencies of defect detections than dynamic displays. The mean detection frequency for each display type is shown in the left portion of Table 1. Overall, there was a decrease in detections over trial blocks. This decrease, however, was due primarily to static displays. Inspectors using dynamic displays maintained an inferior but constant detection rate whereas inspectors using static displays showed superior detection rates over the first 50 defects, but detected fewer of the last 25 defects (Newman-Keuls, $p < .05$). Obviously, the latter result is due to the fact that most inspectors using static displays did not complete the inspection of all the discs in the time allotted.

The main effect of display arrangement was not significant ($p > .05$). Figure 1 shows that the only effect of display arrangement appeared when ordered arrangements were combined with static displays. Under this combination, detections decreased over each block of defects (Newman-Keuls, $p < .05$).

Figure 1 also clearly shows that overall, more defects were detected during the experimental trial than during the practice trial. The improvement, however, is restricted primarily to ordered arrangements when dynamic displays are used and random arrangements when static displays are employed (Newman-Keuls, $p < .05$). In order to test for a possible overall practice effect during the eight inspection trials, the data were grouped into the order of presentation of conditions (four levels) and analyzed by means of a repeated-measures analysis of variance. No significant overall practice effects resulted from analyses of either detection or false alarm data ($p > .05$). Consequently, any inspection improvement seems to be restricted to particular combinations of display types and arrangements rather than overall practice.

DISCUSSION

Structuring according to display arrangement seems to have little effect on inspector performance. In fact, many *S*s after completion of the experiment stated they were not aware that discs had been displayed in both a random and ordered fashion. This result is possibly due to the close proximity of inspection items in both the ordered and random arrangement.

Surprisingly, any structuring by display type

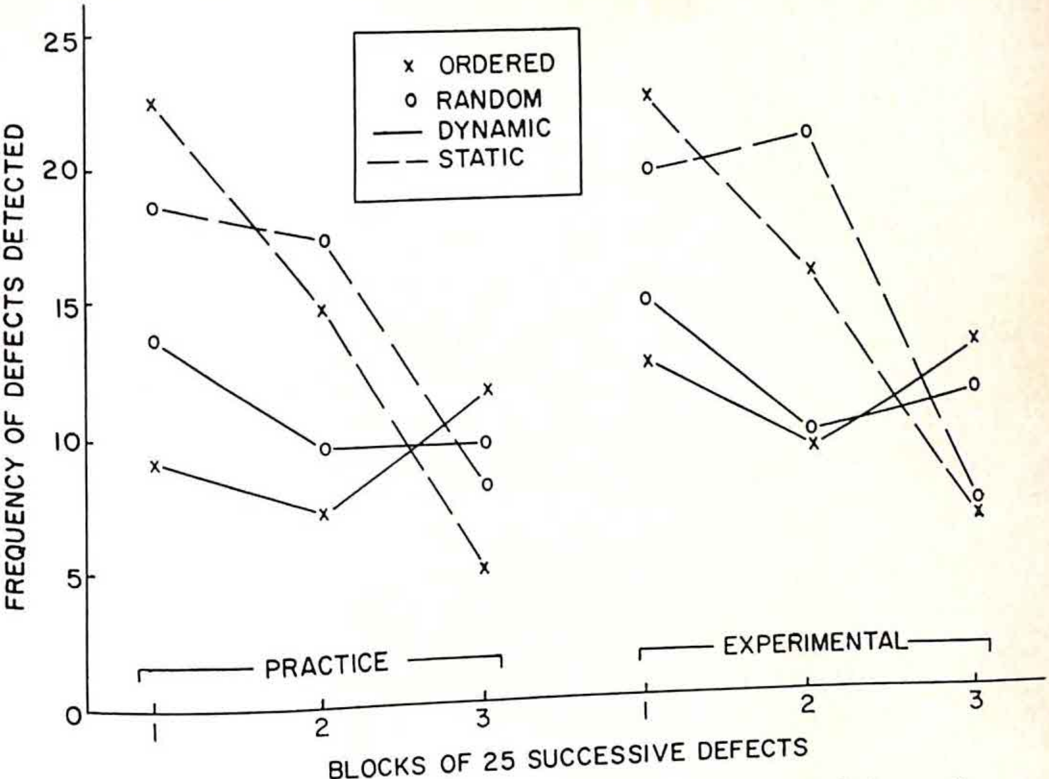


Fig. 1. Frequency of defect detection in successive thirds of the total task during practice and experimental trials as a function of display type and arrangement.

seems to hinder rather than enhance inspector performance. When inspectors are forced externally to pace their inspections at a rapid rate, they do so at the expense of increased errors and fewer detections overall.

One explanation of these data is that inspectors emphasized accuracy rather than speed. Implicit in inspection tasks is the criterion of accuracy. Consequently, when not paced, inspectors will perform at a much slower rate, but with more accuracy.

Many inspection tasks, however, demand a fast rate of inspection and operate on a paced display (e.g., assembly line inspections). The major result of this study suggests that inspectors will, indeed, pace their inspections according to the display presentation rate, but the overall result may be inferior inspection performance when compared to alternative inspector-paced displays. The implication of these results for the design of inspection tasks suggests an inspector-paced display seems to be more beneficial than an externally paced display when rapid scanning rates are required. In addition, the results suggest a third scanning procedure which needs to be examined—intermittent pacing. An inspector being paced intermittently would be presented a static group or chunk of items to be inspected at his own rate while each group is being externally paced. When external pacing is required, such a scanning procedure might prove advantageous.

REFERENCES

Badalamente, R. V., & Ayoub, M. M. A behavioral analysis of an assembly line inspection task. *Human Factors*, 1969, 11, 339–352.

Harris, D. H. Effects of equipment complexity on inspection performance. *Journal of Applied Psychology*, 1966, **50**, 236–237.

Harris, D. H. Effect of defect rate on inspection accuracy. *Journal of Applied Psychology*, 1968, **52**, 377–379.

Harris, D. H. The nature of industrial inspection. *Human Factors*, 1969, 11, 139–148.

Wallack, P. M., & Adams, S. K. The utility of signal-detection theory in the analysis of industrial inspector accuracy. *AIIE Transactions*, 1969, 1, 33–44.

(Received November 24, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 126–134

PUPIL DILATION AND ITS RELATIONSHIP TO INFORMATION PROCESSING DURING READING AND LISTENING[1]

RONALD P. CARVER[2]

American Institutes for Research, Washington, D. C.

Pupil diameter has been advanced as an index of information-processing load. The question of whether pupil diameter indicates information-processing load during reading was investigated in three experiments involving 24 college students reading passages of varying difficulty. A TV camera and monitor, together with a video recorder, were used to measure the diameter of the pupil under a reading condition and three control conditions. No evidence was obtained which would suggest that the pupil dilates while reading. The generality of these results was investigated by measuring the pupil size of 20 college students while listening to tape recordings of the reading passages. No evidence was obtained which suggested that the pupil dilates while listening. These results suggest that pupil size cannot be used as an objective indication of whether or not an individual is, in fact, processing the information contained in connected discourse while he is supposedly engaged in reading or listening activities. Also discussed was the possibility that dilations occurred but were too small to be detected with the methods used.

Research has suggested that pupil diameter indicates the degree of mental effort or mental activity involved in such tasks as problem solving (Hess & Plot, 1964), word imagery (Paivio & Simpson, 1966), short-term memory (Kahneman & Beatty, 1966), and tone discrimination (Kahneman & Beatty, 1967). Beatty and Kahneman (1966) further contend that the pupil diameter is a valid measure of the momentary state of mental effort. Stated differently, Kahneman and Beatty (1967) have referred to the validity of the pupillary measure as an index of information-processing load. They explain this concept by reference to an analogy with the aggregate demand for power (total amperage) in an electrical system.

It would be helpful in research on learning from connected discourse if there existed an objective indicator of the moment-to-moment information-processing activity of individuals. Presently, there is no method for either studying or controlling for these variables. Psychometric tests are employed to give post hoc estimates of these activities, but the imprecision of these methods make advancement in theory and practice extremely difficult. The major purpose of the following research was to investigate pupil dilation as a possible indicator of the information-processing activities of individuals while reading and listening.

A minor purpose of the research was to investigate whether the difficulty of the material would be related to pupil size. Payne, Parry, and Harasymiw (1968) studied pupil dilation as a measure of the difficulty of multiple-choice test questions. They stated that pupil dilation is a sensitive and apparently reliable indicator of internal information processing, and warrants further study. They interpreted their results as suggesting that pupillary dilation is an indicator of item difficulty but is not as good as another indicator they studied (i.e., the time taken to complete the item). These results do suggest, however, that the difficulty of the reading or listening task may also affect the pupillary dilation.

In general, the purpose of the following research was to investigate pupil dilation during the processing and storage of the information contained in connected discourse. At the outset, the effect of reading upon pupillary dilation was investigated in a series of experiments and then the mode was shifted to listening for the final experiment.

---

[1] This research was supported by the United States Office of Education, Region III, under Grant No. OE6-3-9-090009-0022(010).

This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Ronald P. Carver, American Institutes for Research, 8555 Sixteenth Street, Silver Spring, Maryland 20910.

EXPERIMENT I

The major purpose of Exp. I was to determine the effect of reading upon pupil dilation. A minor purpose was to investigate the effect of the difficulty of the reading material upon pupil dilation. Since dilation is always in relation to a control condition, the selection of an appropriate control was critical. Three control conditions were used: fixation upon a point approximately 11 in. away, fixation upon each of a series of numbers, and fixation upon each of the words in a reading passage which had been turned upside down.

Method

Subjects. Eight college student volunteers were paid for their 1 hr. of participation in the experiment.

Equipment. The pupil was videotaped, using a system which included a Panasonic TV camera, Model WV-220*P*; an adjustable extension tube on a Makro-kilar lens (1:2.8/90) manufactured by Heinz Kilfitt Muchen; a 1-in. video recorder, Panasonic Video Master, 2100 SD; two adjustable desk lamps (two 15-w. fluorescent tubes in each lamp); a head positioning apparatus, which provided a stationary forehead rest and an adjustable chin rest; and a clamp device for holding the stimulus cards.

The TV camera was mounted on a $22\frac{1}{2} \times 36$ in. horizontal surface, 37 in. above floor level. The camera was mounted so that its horizontal position was in the same horizontal plane as the eyes and at a 25° angle to a line perpendicular to the left eye. The eye and camera lens were approximately 13 in. above the mounting surface and the camera was focused upon the left eye. The fluorescent lamps were positioned on each side of the head so that each lamp was in a vertical position 6 in. from the eye and at a 45° angle from the line of sight.

Materials. The stimulus materials were 5×8 in. white index cards which had been cut to 5×5 in. The number card was a stimulus card with the numbers 1–9 typed on the card. The 1 was on the left part of the card, the 2, top middle, the 3, top right, the 4, middle left, and so on. The numbers thus provided fixation points on the card for eight border positions and one center position.

There were eight different reading cards, each containing one 150-word reading passage. The passages were selected from the set of 36 passages originally scaled for difficulty by Miller and Coleman (1967). They were selected to represent four levels of difficulty. Recently, Aquino (1969) has further studied the difficulty level of these passages and has published the 36 paragraphs in their entirety. Their order of presentation in Aquino's article are used as an identification number when referring to specific passages. The eight paragraphs (two at each difficulty level) used in the study are shown in Table 1 represented by their Aquino identification numbers and corresponding number, and letter identification used in the present study.

TABLE 1

AQUINO IDENTIFICATION NUMBERS AND IDENTIFYING CODE DESIGNATIONS FOR THE EIGHT PASSAGES

Difficulty	Set 1		Set 2	
	Aquino no.	Identification code	Aquino no.	Identification code
Very easy	13	1A	10	1B
Moderately easy	22	2A	25	2B
Moderately difficult	27	3A	32	3B
Very difficult	33	4A	36	4B

The comprehension tests for these reading passages were Cloze tests (see Taylor, 1953). Ten words were arbitrarily selected in each passage as test items and these words were deleted when the passage was retyped as a test. The items were all selected from the lexical words (i.e., nouns, adjectives, adverbs, and verbs) rather than from the function words (i.e., articles, conjunctions, prepositions, and pronouns). The items were also selected to be approximately equally spaced throughout the passage.

Instructions. The following instructions were read to *S* at the beginning of the experimental session: "This is an experiment to investigate the function of the eyes when reading. You will be asked to read some short paragraphs and then take a sentence completion test on the paragraphs."

Procedure. Upon entering the laboratory room, *S* was: (*a*) instructed to sit in front of the experimental apparatus; (*b*) given the instructions concerning the nature of the experiment; (*c*) asked to place his chin in a chin rest, and (*d*) asked to fixate upon the number 5 on the center of the card in front of him. If necessary, *E* then adjusted: (*a*) the chin rest; (*b*) the height of the chair; or (*c*) the focus of the television camera. The *E* focused on *S*'s left eye so that its image was in the center of the TV monitor.

After adjustments were made, *S* was asked to move away from the apparatus and was given a verbal preview of the practice procedures to follow. The *S* was then asked to return to the chin rest, and the focus and position of the eye were readjusted if necessary. The *S* was instructed to close his eyes for approximately 3 sec. The *E* then activated a stopwatch, instructed *S* to open his eyes, and started the videotape simultaneously. The *S* fixated upon the number 5 in the center of the card (base-line fixation control condition, BFC).

At the end of 20 sec., *S* was instructed to close his eyes and, after 5 sec., was instructed to open them and look at each of the numbers 1–9 on the card. The *S* was instructed to continue looking at each of the numbers 1–9 (i.e., when 9 was reached, *S* was to return to 1 and start again (eye-movement control condition, EMC).

At the end of 20 sec., *S* was signaled to close his eyes. The *E* then removed the number card from the viewing

TABLE 2

PASSAGE PRESENTATION ORDER FOR A SET OF FOUR *S*s

S	Presentation order							
	1	2	3	4	5	6	7	8
1	1A	2A	3A	4A	1B	2B	3B	4B
2	4A	1A	2A	3A	4B	1B	2B	3B
3	3A	4A	1A	2A	3B	4B	1B	2B
4	2A	3A	4A	1A	2B	3B	4B	1B

position. The *S* was then instructed to open his eyes and begin reading immediately (condition, READ).

When *S* finished the passage, he was instructed to close his eyes. The *E* recorded the reading time with an electric timer. With *S*'s eyes closed, *E* turned the passage card upside-down and then instructed *S* to open his eyes and look at each word on the card as though he were reading (upside-down looking control condition, ULC). That is, *S* was to start in the upper left corner, look at each upside-down word from left to right and continue down each line of the card until the last word was looked at in the lower right-hand corner of the card. The *S* was asked to refrain from trying to recognize the meaning of the upside-down words while he was looking at them.

The *S* was instructed to close his eyes when he finished reading. The *E* used a timer to record the duration of this condition. If *E* noticed on the TV monitor that *S* tended to skip over words, he was instructed to decrease his rate of looking at the upside-down words so that he was fixating upon each word. These procedures produced looking-time durations which were approximately equal to the corresponding reading-time durations.

After the ULC condition, *S* was asked to move away from the apparatus and take a seat at a desk where the test on the passage was administered. The *S* was told that the passages had four levels of difficulty: very difficult, moderately difficult, moderately easy, and very easy. The *S* was told the difficulty level of his particular practice passage so that he would have some idea of what to expect for the experimental passages. The *S* was given 3 min. to complete the test. (Most individuals finished each test in less than 2 min.) Then, the entire procedure was reviewed and *S* was informed that for the experiment he would repeat the same procedural cycle for seven more passages.

Design. Table 2 contains the general design for the study. The *S*s were tested individually and were assigned sequentially to the four Latin-square passage orders. The first passage presented was always the practice passage, and the following seven were experimental passages. However, in the data analysis, Passages 2–4 were also considered by *E* as practice passages to allow anxiety and the learning to learn effect to stabilize.

Measures. The *S*s were scheduled at 2-hr. intervals and the hour between *S*s was used to record the data from the videotape. Three data points were sampled from each of the four conditions (BFC, EMC, READ, and ULC) for the four paragraphs (1B, 2B, 3B, and 4B). Thus, a total of 48 pupil diameters was measured for each *S*.

The sampling procedure for each condition was as follows: BFC–Sample 1 was taken 10 sec. from the beginning of the condition, Sample 2 was taken approximately 2 sec. later, and Sample 3 was taken approximately 2 sec. after Sample 2. EMC—Sample 1 was taken from the first fixation on the number 5 following the 10-sec. point in the condition interval (i.e., *S* completed a cycle of looking at 6, 7, 8, 9, 1, 2, 3, and 4 before the next sample was taken); Sample 2 was taken from the next fixation on the number 5, and Sample 3 was taken from the following fixation on number 5 after Sample 2. READ—Sample 1 was taken from the first fixation in the center of the monitor screen following the 10-sec. point in the condition interval, Sample 2 was taken when *S* had finished the line he was reading for Sample 1 and had again reached the center of the screen while reading the following line, and Sample 3 was like Sample 2 except the measurement was taken from the following line. ULC—Samples 1–3 were taken in a manner exactly like those in the READ condition.

The pupil diameter was measured directly from the stop-action image on the monitor screen. The measurement was accomplished by affixing a transparent grid (10 squares per inch) to the monitor screen. The horizontal coordinate which was tangent to the upper edge of the pupil was observed and recorded by *E*; likewise, the horizontal coordinate which was tangent to the lower edge of the pupil was also observed and recorded by *E*. Thus, the pupil diameter was always the vertical diameter and this value was calculated by subtracting the value recorded for the lower edge from the upper edge.

If any planned measurement coincided with or immediately followed an eye blink, the videotape was reversed in slow motion, and the measurement was taken immediately preceding the eye blink to eliminate the unwanted dilation immediately following the eye blink. Also, all measurements were taken during fixations since the image was blurred during eye movements.

All pupil diameters were therefore recorded in inches as they appeared on the monitor. These diameters varied between individuals from approximately 2.5–5.0 in. The actual size of the pupil in millimeters was not estimated, since the distance between the eye and the camera varied with the contour of *S*'s head; the focus, therefore, varied slightly between individuals. However, by focusing upon circles of known diameter placed in a normal position of the eye, the magnification of the system was estimated. A diameter change of .1 in. on the monitor screen was thus estimated to equal .00565 in., or .14 mm., on the actual eye. The magnification on the monitor was therefore approximately 18 times the actual size of the pupil for most individuals. It was impossible to estimate the horizontal tangent to the pupil closer than the nearest .1 in. and, therefore, this was the precision of the measurements—.1 in. on the monitor or .14 mm. of pupil diameter.

After each of the three samples had been taken for each passage condition, the median value was selected to represent the condition for additional data analyses (i.e., for each individual reading each passage there was one pupil diameter for each of the four conditions in the data analysis). Since slight changes in the head position or movement in the apparatus often necessitated refocusing at the beginning of each passage trial, pupil diameters could not be directly compared between individuals or between passage conditions within individuals. In order to make these comparisons, pupil diameters under each of the three conditions—EMC, READ, and ULC—were adjusted for base-line differences by subtracting the pupil diameter under the BFC condition.

Secondary to the pupil diameter measurements were the passage reading time data and the test scores on the passages. These data were analyzed solely to ascertain whether the difficulty of the reading passages was reflected in reading durations and comprehension scores. The times recorded for looking at the upside-down words were not analyzed. These durations were recorded only to provide additional immediate feedback to *E* relevant to the similarity between the eye movements under this condition and the eye movements during the reading condition.

Results and Discussion

The mean reading times for the four passages increased with the difficulty level of the passage (1B, .63 min.; 2B, .78 min.; 3B, .79 min.; and 4B, 1.08 min.), and the mean number correct on the comprehension tests decreased with the difficulty level (1B, 9.6; 2B, 6.9; 3B, 5.4; 4B, 2.4). It therefore appeared that the difficulty variable had been adequately manipulated.

Table 3 contains the mean pupil dilation for the four passages under the two control and one experimental conditions. There were only four mean dilations which were greater than the accuracy of the measurement, .10 in., and the largest dilation was only .20 in. These differences are quite small in relation to the size of the dilation, approximately .50 mm., reported by Kahneman and Beatty (1966). A .50-mm. dilation would result in a change of approximately .36 in. on the TV monitor in the present study.

At this point in the data analysis, a contaminating variable was discovered. Although most of the measurements were made in the center of the screen, some were not. A correlation was discovered between the position of the pupil on the screen and the pupil size. That is, the screen was distorted so that the pupil increased slightly in diameter as it changed in position on the screen from bottom to top. During the reading of the more difficult paragraphs, the individuals had not progressed as far down the screen during the first 10 sec. of the READ condition as they had during the less difficult passages. Thus, these latter pupil dilations tend to be larger.

TABLE 3

MEAN PUPIL DILATION IN RELATION TO BASE LINE FOR THREE CONDITIONS AND FOUR READING PASSAGES

Passage	Conditions		
	EMC (Control)	READ (Experimental)	ULC (Control)
1B	.04	.05	.03
2B	.06	.14	.08
3B	.01	.11	.02
4B	.11	.20	.09
$\bar{X}$	.06	.12	.05

Note.—N = 8; Mean values are in inches of dilation as measured on a TV monitor which had magnified the pupil approximately 18 times.

A correction factor was formulated by using a test circle of consistent diameter. When the correction factor was applied to the data, the differences in Table 2 became less marked. These corrected values are presented in Table 4. Notice that the largest dilation was only .17 in. and that there appeared to be no trend for the dilation to increase as the difficulty of the passage increased. Furthermore, the mean of the mean differences for each of the three conditions were all approximately equal. Thus, there appeared to be little evidence to suggest that the pupil dilated while the individuals were processing information during the reading of the paragraphs.

It did not appear fruitful to try to replicate the results in Exp. 1 under precisely the same experimental conditions.

EXPERIMENT II

In order to be relatively confident that the motive-incentive conditions were not a major variable affecting the pupil dilation, the instructions in the following study were modified in a manner designed to increase motivation. Furthermore, the measurement procedures were modified slightly to further control for the screen distortion.

TABLE 4

MEAN PUPIL DILATION IN RELATION TO BASE LINE FOR THREE CONDITIONS AND FOUR READING PASSAGES AFTER CORRECTION FOR SCREEN DISTORTION

Passage	Conditions		
	EMC (Control)	READ (Experimental)	ULC (Control)
1B	.07	.07	.07
2B	.09	.17	.14
3B	.05	.09	.06
4B	.11	.16	.06
$\bar{X}$	.10	.12	.08

Note.—$N = 8$; Mean values are in inches of dilation as measured on a TV monitor which had magnified the pupil approximately 18 times.

Method

The method was exactly the same as in Study 1 except for the following three changes:

1. The instructions read to each *S* were changed to increase motivation by providing *S* with knowledge of results and by indicating that *S*'s performance might be categorized by *E* as bad. The instructions were: "This experiment is designed to investigate the function of the eyes during reading. We are interested in the differences in function between good and poor readers. You will read a number of short paragraphs and then take sentence completion tests on these paragraphs. I will not comment on how well you did each time, but I will grade each test immediately after you have finished so that you will know how may blanks you filled in correctly."

2. The *E* marked the incorrect responses in the presence of *S* immediately after he had finished each of the eight tests.

3. The measurement procedures were modified so that fewer measurements were made in positions other than the exact center of the monitor screen. That is, if *S* moved his head slightly under condition EMC, so that his pupil was in the center of the screen when he was fixating, for example, on 6 instead of 5, then the center fixation was used instead of the 5 fixation. Also, under conditions READ and ULC, instead of always selecting the first sample at the 10-sec. point, the first sample was selected for a line that was approximately in the center of the screen. With these modifications it was still imperative that the correction factor, developed in Exp. I, be applied in many instances, since *S*s often changed their head position slightly.

Results and Discussion

Mean reading times for the four passages again increased with the difficulty level of the passage (.75, .83, .95, and 1.33 min.) and mean number correct on the comprehension tests again decreased with the difficulty level (8.8, 6.6, 4.0, and 2.2). Although the mean reading times were higher in Study 2 than in Study 1, the mean comprehension scores were lower. Thus, it appears that the group in Study 2 was probably slightly less able, rather than more motivated. These data suggest that *S*s were probably doing their best in both Study 1 and Study 2.

Table 5 contains the mean pupil dilations for the four passages under the two control and one experimental conditions. Notice that no mean dilation is greater than the accuracy of measurement, .10 in., and the average dilation over all four passages was equal, .05 in., for all three conditions.

These data further confirm the results of Exp. I and provide no evidence that the pupil dilates when individuals are processing the information contained in connected discourse.

It did not appear to be fruitful to replicate these results using the same procedures.

EXPERIMENT III

The major purpose of Exp. III was to further extend the results of Exp. I and II under slightly modified procedures. In this study, *S*s were not allowed to read at a rate of their own choosing, but were given a short fixed amount of time to spend reading. This procedure was used to produce a consistently high level of information processing load for

TABLE 5

MEAN PUPIL DILATION IN RELATION TO BASE LINE FOR THREE CONDITIONS AND FOUR READING PASSAGES

Passage	Conditions		
	EMC (Control)	READ (Experimental)	ULC (Control)
1B	.04	.05	.06
2B	.00	.02	.01
3B	.08	.09	.06
4B	.06	.05	.05
$\bar{X}$	.05	.05	.05

Note.—$N = 8$; Mean values are in inches of dilation as measured on a TV monitor which magnified the pupil approximately 18 times.

all levels of passage difficulty. That is, *S* would be expected to try to process as much information as he could during the time interval, thus keeping the load at a maximum during the entire 20 sec.

Method

The method was exactly like that in Study 2, with the exception that *S*s were given only 20 sec. to read each paragraph and 20 sec. to look at the upside-down words.

Results and Discussion

Mean number correct again decreased with the difficulty level of the passages (7.5, 5.3, 2.6, and 1.2), and, in each of the four comparisons, the scores were lower than those of Studies 1 and 2. Although it was never seriously questioned that *S*s were processing information while they were looking at the words, it was possible that the items could have been answered correctly without reading. Thus, the lower mean scores in Exp. III provide some evidence that *S*s were processing information while they were fixating upon the words under the reading conditions.

TABLE 6

MEAN PUPIL DILATION IN RELATION TO BASE LINE FOR THREE CONDITIONS AND FOUR READING PASSAGES

Passage	Conditions		
	EMC (Control)	READ (Experimental)	ULC (Control)
1B	.08	.21	.15
2B	.11	.22	.20
3B	−.01	.12	.12
4B	.04	.06	.10
$\bar{X}$	.05	.15	.14

Note.—$N = 8$; Mean values are in inches of dilation as measured on a TV monitor which magnified the pupil approximately 18 times.

Table 6 contains the mean pupil dilation in relation to base line for the three passages under the two control and one experimental conditions. Although the mean of the four mean passage dilations under the reading condition, .15 in., was greater than the accuracy of measurement, .10 in., it was essentially equal to the mean dilation for all four passages under the ULC condition, .14 in.

Therefore, the results of Studies 1–3 do not provide evidence which would suggest that the information being processed while reading is associated with a greater pupil size. It is possible that these results cannot be generalized to another mode of processing the information contained in connected discourse. Furthermore, it is possible that some other variable is associated with eye movements which contaminated the data in these studies, such as the measurement procedures. Therefore, it appeared desirable to investigate the effect of listening upon pupil dilation. Listening is another mode of processing the information contained in connected discourse. Listening has the advantage of allowing pupil dilation to be measured during the information-processing activity without accompanying eye movements, as in reading.

EXPERIMENT IV

The purpose of the following study was to investigate the effect of listening upon pupil dilation to determine if the results of Studies 1, 2, and 3 could be generalized beyond the reading mode of processing the information contained in connected discourse.

Method

The method was exactly the same as that in Exp. II with the following exceptions:

1. There were 20 *S*s instead of 8.
2. The eight passages were tape recorded by a professional announcer and *S* simply fixated upon the 5 on the number card while he listened to the passages as they were being played. Thus, the previous READ condition was replaced by a listening condition.
3. The passage order was always the same for each *S*—1A, 2A, 3A, 4A, 1B, 2B, 3B, and 4B.
4. There was only one control condition for each passage and that was the 20-sec. base-line condition, BFC. The other two control conditions were necessary for reading only.
5. There were five samples taken from each of the two conditions—BFC and listening. For each condition a measurement was taken at the instant that the eyelid opened to a position above the pupil, and at 5-sec. intervals thereafter to 20 sec. (i.e., 0, 5, 10, 15, and 20-sec. points). The 0-sec. point was measured to provide an indication of the precision of the measurement procedures.
6. The dependent variable analyzed was the pupil diameter during the listening sample minus the diameter during the corresponding control sample. For example, the pupil dilation at the 5-sec. point for the

TABLE 7

MEAN PUPIL DILATION IN RELATION TO BASE LINE FOR THREE SAMPLING POINTS AND FOUR LISTENING PASSAGES

Passage	Sample points (in sec.)				$\bar{X}$
	5	10	15	20	
1B	.02	.07	.00	.03	.03
2B	.12[a]	.04	.01	.01	.04
3B	.04	.00	.00	.00	.01
4B	.04	−.04	.01	.06[a]	.02
$\bar{X}$	.05	.02	.00	.02	

Note.—$N = 20$.
[a] t-test significant beyond the .05 level.

control condition was subtracted from the diameter at the 5-sec. point for the listening condition. This measure of dilation was calculated for each of the four sample points (5, 10, 15, 20) and the four passages (1B, 2B, 3B, and 4B).

Results and Discussion

Mean number correct for the four passages were approximately equal to the means in Studies 1 and 2 (7.8, 6.8, 4.6, and 2.6).

Table 7 contains the mean pupil dilations in relation to the base-line control (BFC) for four sampling positions and the four passages. Notice that only two of the 24 t tests for dilation were statistically significant. These two statistically significant differences were probably chance occurrences, since neither the marginal means for sampling points or levels of difficulty suggest a trend. All the marginals are close to zero, and none are above the accuracy of measurement, .10 in. However, in order to be sure that the size of the mean dilations were not large in relation to their variances, an analysis of variance was conducted. An ABS design (see Lindquist, 1956) was applied with the A variable representing the four sample points and the B variable representing the four passages varying in difficulty level. Neither the F representing the main effects of A and B, nor the F representing their interaction approached statistical significance at the .05 level.

Thus, it appears that there is no evidence which would suggest that the pupil dilates while processing the information contained in aurally presented connected discourse.

In order to demonstrate that the procedures, techniques, and equipment were probably sufficiently sensitive to dilation had it occured, the mean pupil diameter was plotted (see Figure 1) as a function of the sampling point for both the experimental and control conditions. These values represent an averaging over all four of the listening passages. Notice that there was a large drop in pupil diameter from the 0 point to the other four sampling points. Figure 1 graphically depicts the potential amount of dilation and the precision of measurement procedures as well as the essential equality of pupil size in the control and listening conditions.

In order to estimate the actual size of pupil dilation that these procedures could have detected as statistically significant, further calculations were made. For the data collected under the 5-sec. sample point and the 4B paragraph, the standard deviations under the listening (.48 in.) and control (.57 in.) conditions, together with the correlation between these two conditions ($r = .96$), were used to estimate the mean difference that would have been statistically significant at the .05 level. This mean difference transformed into actual pupil size was .09 mm. That is,

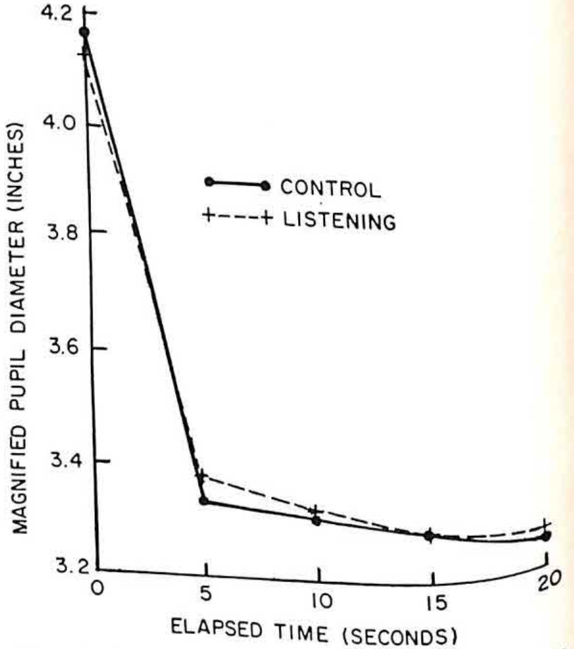


FIG. 1. Mean pupil diameter as a function of the time elapsed from the opening of the eye during a listening condition and a nonlistening control position ($N = 20$).

it was estimated that a pupil dilation of around .09 mm. would be detected as statistically significant in Exp. IV.

DISCUSSION

The results of Studies 1, 2, 3, and 4 suggest that the pupil does not dilate while reading or when *S* is listening to connected discourse. That is, it does not appear that pupil size can be used as an objective indication of whether or not an individual is, in fact, processing the information contained in connected discourse while supposedly engaged in reading or listening activities. What appeared at the outset to be a potential means of controlling and investigating the moment-to-moment changes in information processing and storage during reading and listening has not provided empirical data which substantiated this potential.

The data collected were not designed to provide definitive evidence with regard to the validity of pupil dilation as a measure of information-processing load. However, these data are indirectly relevant to this concept. It appears certain that *S*s in the four studies were processing information during the experimental conditions. Furthermore, it seems reasonable to infer that the information-processing load was greater under the reading and listening conditions than under the control conditions. Thus, an explanation for the failure to find dilation would seem to be in order.

One explanation for these results could be that the information-processing load was minimal and the research techniques were not precise enough to detect the corresponding small dilation. This explanation is highly plausible. It seems reasonable that the load during the processing of information during reading or listening may be small in relation to the load during the storage of a seven-digit number (see Kahneman & Beatty, 1966). Since the conduct of the research reported here, the results of other pupil-dilation experiments have become available, for example, Kahneman & Peavler, 1969, who report pupil dilations of very low orders of magnitude (around .06 mm.). Kahneman and Peavler have discussed an alternative causal factor in their data, that is, anxiety or emotion. It is quite possible that a more adequate theoretical explanation for pupillary dilation is not level of information-processing load, but the anxiety that many times accompanies the higher levels of information processing. Kahneman, Peavler, and Onuska (1968) report dilations up to .30 and .40 mm. for college students during the storage of a four-digit number. Surely, the degree of information processing involved in this task would not be greater than the reading and listening activities involved in Exp. I–IV. However, it is conceivable that the anxiety level when storing the four-digit numbers was higher than that in the present studies, since it would be embarrassing to fail to remember a four-digit number or to fail to report the numbers on cue.

Kahneman and Peavler (1969) have interpreted the results of recent pupil-dilation research as rendering its validity practically unassailable when used as a measure of mental effort or processing load. Yet, there appears to be no research reported which was explicitly designed to separate anxiety or emotion effects from information-processing effects. This situation, taken together with the present failure to find a dilation during reading and listening which was as large as the dilation that accompanies the processing of a four-digit number, seems to make it reasonable to withhold judgment concerning validity. It appears that the alternative hypothesis—that the anxiety or emotion which often accompanies higher levels of information processing is responsible for pupil dilation—still deserves attention in future research.

Another problem involved in the comparison of these results with previous research data is the use of the terms, information processing, processing load, mental effort, anxiety, and emotion. It is practically impossible to operationally discriminate among these terms, and conceptual discriminations are also difficult. It seems that another explanation for the failure to find dilation may involve a failure to be precise in the discrimination between information-processing load and processing load. That is, mental effort may be interchangeable with processing load as Kahneman and Peavler (1969) have suggested, but information-processing load may not be interchangeable with these two concepts as the discussion of Kahneman and Beatty (1967) would suggest. As an example of this distinction, it seems reasonable that experimental conditions may make the level of processing load (or mental effort, or anxiety, or emotion) higher during the storing of a four-digit number as compared to the level during reading and listening. However, it does not seem reasonable that the information-processing load during the storage of a four-digit number would be higher than that during reading and listening. Before it is possible to be confident about the validity of pupil dilation as an index of a theoretical construct, considerable progress needs to be made in both the conceptual and operational definition of terms.

In the interpretation of the results of Exp. I–IV, it is impossible to know whether small dilations accompanied the information-processing activities or whether none existed. In either event, however, it seems impractical, at this state of knowledge and measurement precision, to use pupil dilation as an indicator of information-processing load during reading and listening.

REFERENCES

AQUINO, M. R. The validity of the Miller-Coleman readability scale. *Reading Research Quarterly*, 1969, **4**, 342–357.

BEATTY, J., & KAHNEMAN, D. Pupillary changes in two memory tasks. *Psychonomic Science*, 1966, **5**(1), 371–372.

HESS, E., & POLT, J. Pupil size in relation to mental activity during simple problem-solving. *Science*, 1964, **140**, 1190–1192.

KAHNEMAN, D., & BEATTY, J. Pupil diameter and load on memory. *Science*, 1966, **154**, 1583–1585.

KAHNEMAN, D., & BEATTY, J. Pupillary responses in a pitch-discrimination task. *Perception & Psychophysics*, 1967, **2**, 101–105.

KAHNEMAN, D., & PEAVLER, W. S. Incentive effects and pupilary changes in association learning. *Journal of Experimental Psychology*, 1969, **79**, 312–318.

KAHNEMAN, D., PEAVLER, W. S., & ONUSKA, L. Effects of verbalization and incentive on the pupil response to mental activity. *Canadian Journal of Psychology*, 1968, **22**, 186–196.

LINDQUIST, E. F. *Design and analysis of experiments*. Boston: Houghton Mifflin, 1956.

MILLER, G. R., & COLEMAN, E. B. A set of thirty-six passages calibrated for complexity. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 851–854.

PAIVIO, A., & SIMPSON, H. The effect of word abstractness and pleasantness on pupil size during an imagery task. *Psychonomic Science*, 1966, **5**(2), 55–56.

PAYNE, D. T., PARRY, M. E., & HARASYMIW, S. J. Percentage of pupillary dilation as a measure of item difficulty. *Perception & Psychophysics*, 1968, **4**, 139–143.

TAYLOR, W. L. "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, 1953, **30**, 415–433.

(Received November 24, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 135–137

VOCATIONAL INTERESTS OF POLITICAL SCIENTISTS [1]

JACK E. ROSSMANN [2]
Macalester College

ORVILLE J. LIPS
Iowa State University

AND DAVID P. CAMPBELL
University of Minnesota

Strong Vocational Interest Blanks were collected from 177 male members of the Midwest Conference of Political Scientists. Their interests were most like those of psychiatrists and psychologists, public administrators and social workers, librarians and musicians, and lawyers and author-journalists. When compared with sociologists, political scientists had interests less like those of applied biological scientists, physical scientists, outdoors and skilled worker occupations, administrators, and businessmen. Similar differences were found between psychologists and political scientists.

As increasing percentages of undergraduate students enroll in graduate schools, more of them may be uncertain about their choice of major. Much of this uncertainty will be due to the difficulty of choice between similar fields. For example, a student may be quite certain that he wants to do further study in the social sciences, but, within that general area, may have trouble deciding between areas such as sociology, psychology, political science, or anthropology. Can a measure of vocational interests such as the Strong Vocational Interest Blank (SVIB) be of help to such students?

The SVIB has demonstrated its usefulness in differentiating among such varied occupational fields as the biological sciences and physical sciences, between art and business, between social service and verbal-linguistic areas. But finer differentiations have been more difficult to make. SVIB data collected from sociologists (Rossmann, Lips, & Campbell, 1967) indicated that while a number of statistically significant differences were found between profiles of sociologists and psychologists, their basic patterns were very similar and also resembled those of a group of graduate students headed toward careers in college teaching (Rossmann & Bentley, 1969).

Thus, some better scheme for differentiating among the interests of people in the academic disciplines should be attempted. Kuder (1963) and Campbell (1966) have expressed concerns about the problem of building scales that discriminate between occupations while retaining occupational similarities. Campbell suggests that since the SVIB is generally used to indicate a general area of interest to the client, rather than a specific occupation, attempts at sharper occupational differentiation may not prove to be particularly fruitful. Kuder, on the other hand, feels that sharpness in differentiation may be lost by the exclusive use of the men-in-general technique.

With the latter point in mind, we have extended the earlier sociologist research by collecting data from other groups of behavioral scientists. The present study presents data from political scientists and focuses both on their interests and the extent to which these interests can be differentiated from two other groups of behavioral scientists—sociologists and psychologists.

METHOD

SVIBs and a brief questionnaire were mailed to 400 members of the Midwest Conference of Political Scientists in January 1967; completed forms were received from 243 (61%). These were screened to eliminate those who had less than 3 yr. experience ($N = 30$), those without PhDs ($N = 23$), or those who reported that they didn't enjoy their work ($N = 3$). This screening was done because earlier experience with other occupations had suggested that some minimum standards of success and satisfaction should be used to increase the homogeneity of the sample. The final sample included 177 political scientists who can be described as PhDs with at least 3 yr. of experience who say they enjoy their work.

Other data describing the sample are included in Table 1. While the respondents were located in 24

---

[1] This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Jack E. Rossmann, Office of Educational Research, Macalester College, St. Paul, Minnesota 55101.

TABLE 1

BACKGROUND DATA ON 177 PHD POLITICAL SCIENTISTS

Data	N	%
Age		
20–30	10	5.7
31–40	56	31.6
41–50	64	36.2
51–60	31	17.5
61 and over	16	9.0
Date of degree		
Prior to 1946	36	20.3
1946–1955	58	32.8
1956–1965	83	46.9
Employing institution		
University	133	75.1
Liberal arts college	30	17.0
State college	12	6.8
Nonacademic	2	1.1
Major job responsibility		
Teaching	153	86.4
Administration	18	10.2
Research	4	2.3
Nonacademic	2	1.1

states ranging from New York to California and Texas to Minnesota, 83% were from the Midwest. In general, they were predominantly members of a teaching faculty in a university setting and had a median age of 43.

RESULTS

The mean SVIB profile of the political scientists indicated that their interests were most similar to those of psychiatrists and psychologists, public administrators and social workers, librarians and musicians, and lawyers and author-journalists. The interests of political scientists were least like those of physical scientists, members of the skilled and semiprofessional occupations, and businessmen.

The responses of the political scientists to each of the 291 items currently being scored on the SVIB were compared with the item responses of Strong's men-in-general group. Percentage response differences of 20% or greater (in any of the three categories: like, indifferent, or dislike) were found on 103 of the 291 items. This exceeds by 29 the analogous number of differentiating items found in the sociologist data (Rossmann et al., 1967). There is little doubt that the interests of political scientists can be differentiated from those of men in general.

Basic Scales

What does it mean conceptually for members of an occupation to have the kind of interest pattern described above? Campbell et al. (1968) have attempted to clarify interpretations of the occupational scales by developing a set of basic interest scales, including in each, items which correlate relatively highly with each other. Norms for these basic scales were developed by scoring one of Strong's early groups (Strong, 1931) which had been retested in 1966–1967 (as adults).

When political scientists are scored on these basic scales, the four highest scales seem to fit quite well the interests that might be expected for a political science professor—an interest in teaching, speaking, writing, and the subject matter of law and politics. Scales falling at the other end of the continuum (mechanical, agriculture, sales, merchandising, technical supervision, and business management) support the previously mentioned rejection of business and technical occupations.

Inspection of individual items lends further support to these findings. Eighty percent or more of the respondents said they would like to be a college professor, an author of a novel, an editor, a foreign correspondent, and a governor of a state. Equally strong was their dislike for any occupation which involved selling (e.g., auto salesman, life insurance salesman, or real estate salesman).

Comparison with Psychologists and Sociologists

To what extent is the preceding description of the interests of political scientists merely a description of the interests of the academic man or at least behavioral scientists in general and to what extent can the interests of political scientists be differentiated from those of other behavioral scientists?

In an attempt to answer this question the mean scale scores of the political scientists were compared with those of the SVIB criterion group of psychologists (Kriedt, 1948) and a group of PhD sociologists (Rossmann et al., 1967). Of the 49 scales scored, differences of five standard score points (one-half standard deviation) or more were found between the political scientists and sociologists on 24 scales. Differences of the same magnitude were found between political scientists and psychologists on 29 scales. On 20 scales, differences of five standard score points or more were found between political scientists and both psychologists and sociologists.

These differences seem to suggest that as compared to sociologists, political scientists have interests less like those of applied biological scientists, physical scientists, outdoors and skilled worker occupations, administrators, and business men. These same differences hold true when contrasting political scientists and psychologists, but in addition, when compared with psychologists, political scientists have interests more like those of librarians and author-journalists.

Additional meaning is given to these differences when these three groups of behavioral scientists are compared on the SVIB basic scales. Of the 22 scales, differences between political scientists and psychologists of 5 or more standard score points were found on 8 scales while only one difference of this magnitude was found between sociologists and political scientists. Political scientists had higher scores than psychologists on three scales (public speaking, writing, and law/politics) while psychologists' scores were higher on science, mathematics, medical service, technical supervision, and mechanical. Thus, the picture of political scientists as being more verbal/linguistic in orientation but considerably less scientific/mechanical than psychologists emerges.

Rank order correlation (rho) between political scientists and psychologists was .76 while the rho for political scientists and sociologists was .96.

DISCUSSION

From the preceding data, it is apparent that sharp differences do exist between the interests of political scientists and those of men in general. The interests of political scientists also seem to differ somewhat from those of psychologists and sociologists (particularly the former), but an examination of the comparative mean profiles points up the difficulty of translating these differences into practical usefulness for a counselor or individual student. If differences which can be helpful to a prospective graduate student are to be found, the development of a behavioral science men-in-general group may be necessary.

REFERENCES

CAMPBELL, D. P. *Manual for the Strong Vocational Interest Blank*. Stanford: Stanford University Press, 1966.

CAMPBELL, D. P., BORGEN, F. H., EASTES, S. H., JOHANSSON, C. B., & PETERSON, R. A. A set of basic scales for the Strong Vocational Interest Blank for men. *Journal of Applied Psychology*, 1968, **52** (6, Pt. 2).

KRIEDT, P. K. Vocational interests of psychologists. *Journal of Applied Psychology*, 1948, **32**, 482–488.

KUDER, G. F. A rationale for evaluating interest. *Education and Psychological Measurement*, 1963, **23**, 3–10.

ROSSMANN, J. E., & BENTLEY, J. C. Vocational interests of future college teachers. *Vocational Guidance Quarterly*, 1969, **17**, 206–211.

ROSSMANN, J. E., LIPS, O., & CAMPBELL, D. P. Vocational interests of sociologists. *Journal of Counseling Psychology*, 1967, **14**, 497–502.

STRONG, E. K., JR. *Change of interests with age*. Stanford: Stanford University Press, 1931.

(Received December 3, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 138–142

EFFECTS OF TEMPORAL PLACEMENT OF UNFAVORABLE INFORMATION ON DECISION MAKING DURING THE SELECTION INTERVIEW

ROGER N. BLAKENEY[1] AND JOHN F. MACNAUGHTON

University of Houston

One hundred and twenty-six evening school management students rated an applicant for employment after listening to standardized audio tapes of a simulated interview in which the location and content of one piece of unfavorable information was systematically varied. An 18-cell analysis of variance design which counterbalanced the temporal placement of the unfavorable information, the content area in which it occurred, as well as the order of the content areas was used. Dogmatism data were analyzed with multiple-regression techniques. The temporal placement effect was significant at the .01 level. None of the other effects were significant. It was concluded that the temporal placement of the unfavorable information accounted for a statistically significant percentage of the variance in the ratings.

For years, the inadequacies of the decision made during the selection interview have been discussed. However, only in recent years has research into the decision-making process characteristic of this type of interview begun.

Springbett (1958) varied the order of presentation of three kinds of information in the selection interview. He found that the first rating, whether based on the application form, personal appearance of the candidate, or the interview was significantly related to the final rating. Webster (1964) applied the "gating" phenomenon (Bruner, 1957) to provide one possible description of the perceptual process in an interview. Bruner uses the concept of gating to refer to the latter stages of decision making during which an observer's openness to additional cues is drastically reduced. Webster concluded that if favorable data come early in the interview and unfavorable data come late, then there is a tendency toward a favorable decision (accept). This phenomenon should occur because the perceptual process centers around the conformation of the early favorable impression. The unfavorable data should then tend to be disregarded. The reverse is also true; if unfavorable data come early and favorable data late in the interview, then an unfavorable decision (reject) should tend to occur. However, if both occur at the same time, then the effects should cancel each other and a marginal decision should result.

Bolster and Springbett (1961) developed interview protocols using favorable and unfavorable items of information. Interviewers were given a test score to establish a favorable, neutral, or unfavorable set toward the applicant. Then they were asked to read the first item of information on the protocol and rate the applicant, after which they read the second item and rated again, and so on. Their results confirmed the presence of a primacy effect.

Crowell (1961) argued that forcing intermediate ratings, as Springbett (1958) had, may affect the final decision. She found that final ratings were significantly changed by the interpolation of intermediate ratings before the final decisions were recorded even when the content of the interview remained constant.

Wheeler (1969) found dogmatism, as measured by the Dogmatism scale (Rokeach, 1960), to be a major source of variability in the rating of items of interview information. He found that interviewers who obtained high scores on the Dogmatism scale gave significantly less favorable ratings than did interviewers whose scores on the Dogmatism scale were low.

---

[1] Requests for reprints should be sent to Roger N. Blakeney, Department of Behavioral Management Science, University of Houston, Cullen Boulevard, Houston, Texas 77004.

The primary purpose of the present study was to determine the effect of certain variables—and their interrelationships—on the decision-making process in the selection interview. The major focus was on the temporal placement of unfavorable information in the interview. A secondary focus was on the content area or subject matter (work history, educational background, personal-social) in which the unfavorable information was embedded. A third concern was the dogmatism of the rater as measured by the Dogmatism (D) scale.

METHOD

Subjects

The *S*s were 126 evening students in Behavioral Management Science classes of the College of Business Administration, University of Houston. They ranged in classification from juniors to graduate students. The *S*s were volunteers, male, and aged 21–50 with a median age of 25. They had at least 2 yr. of full-time job experience. None of the *S*s had had experience as a personnel interviewer.

Taped Interview

The stimulus to which *S*s responded was a tape-recorded interview in which the temporal placement and nature of a piece of unfavorable information was systematically varied. The design resulted in 18 such interviews. Each recording began with the same brief introduction (approximately 2 min. long) to establish a common set. The remainder of the interview was divided into three segments each 6–6½ min. in length. Each segment dealt with a different content area: work history, academic history, and personal-social background. These segments were followed by a common closing remark by the interviewer which, in a real interview, would have led into the selling of the company. All of the interviews, which were simulated campus recruitment interviews, used the same interviewee and interviewer. They were, in fact, identical except for the order of the three major content areas and the substitution of the unfavorable for the favorable bits of information, since they were edited or pieced together from the same master tape.

Three pieces of unfavorable information, one from each content area, were used. These were all 1½–2 min. long. The unfavorable pieces of information dealt with the following topics: (*a*) work history (W), not liking supervisory responsibility; (*b*) academic history (A), too many extra-curricular activities resulting in low grades; and (*c*) personal-social (PS), inability to get along with people. In each of the 18 interviews, one of these pieces of unfavorable information replaced a piece of favorable information of comparable length and dealing with the same content.

The order in which the three segments occurred was varied systematically to produce all six possible orders, which are as follows:

A	W	PS	PS	A	W
W	A	A	W	PS	PS
PS	PS	W	A	W	A

Then, for each six orders of the three content areas, one of the unfavorable pieces of information was placed in one of three possible temporal locations. This produced the 18 counterbalanced placements (u = unfavorable, f = favorable):

Au	Af	Af	Wu	Wf	Wf
Wf	Wu	Wf	Af	Au	Af
PSf	PSf	PSu	PSf	PSf	PSu

PSu	PSf	PSf	PSu	PSf	PSf
Af	Au	Af	Wf	Wu	Wf
Wf	Wf	Wu	Af	Af	Au

Au	Af	Af	Wu	Wf	Wf
PSf	PSu	PSf	PSf	PSu	PSf
Wf	Wf	Wu	Af	Af	Au

Thus, only one item of unfavorable information occurred during each interview, replacing comparable favorable information. The interviews were, then, identical except that each contained one piece of unfavorable information, the placement of which was varied in counterbalanced order by content area and temporal placement.

Procedure

The 126 *S*s were randomly arranged into 18 sets of 7 each. The number of *S*s listening to a given tape at one time varied from two to four. One of the 18 tapes was played for each set of *S*s. They were instructed to listen to the tape in order to evaluate the interviewee as an applicant for a "general management trainee position in a large industrial corporation." The *S*s were given a rating form on which they rated the applicant on six characteristics and on overall acceptability. The six characteristics—each defined—were ambition, creativity, perseverance, decisiveness, social adjustment, and maturity. However, only the overall rating was used in the analysis.

The *S*s were also given the D scale either before or shortly after they made the ratings.

Hypotheses

The hypotheses of this investigation were as follows:

1. The earlier the unfavorable information comes in the interview, the less favorable the ratings of the interviewee will be; more specifically:

(*a*) When the unfavorable information comes during the first third of the interview, the ratings of the applicant will be the lowest; (*b*) when the unfavorable information comes during the second third, the ratings will be intermediate; (*c*) when the unfavorable information comes during the final third, the ratings will be the highest.

2. The ratings will be lowest where the unfavorable

TABLE 1

MEAN RATINGS OF TAPED INTERVIEWS ARRANGED BY TEMPORAL PLACEMENT OF UNFAVORABLE INFORMATION × CONTENT AREA ORDER

Temporal placement	Content area order						Row $\bar{X}$
	A PS W	A W PS	PS A W	PS W A	W PS A	W A PS	
First third	Tape 5 3.57	Tape 6 3.14	Tape 11 3.43	Tape 12 3.43	Tape 17 3.14	Tape 18 3.00	3.28[a]
Second third	Tape 3 3.29	Tape 2 2.86	Tape 9 3.57	Tape 8 4.14	Tape 15 3.29	Tape 14 3.29	3.40[a]
Final third	Tape 1 4.43	Tape 4 3.43	Tape 7 4.14	Tape 10 3.86	Tape 13 4.14	Tape 16 4.29	4.05
Column $\bar{X}$	3.76	3.15	3.71	3.81	3.52	3.52	3.58

Note.—Rating scale of 1–7. Content area abbreviations: A = academic history, PS = personal-social, W = work history.
[a] $t = .223$ for comparison between mean for first third and mean for second third.

information is embedded in the social-personal content area and comes during the first third of the interview (on the assumption that the social-personal area was subject to more interviewer bias).

3. There will be a negative correlation between D-scale scores and the degree of favorability of the interviewer's ratings.

4. For the high-D *S*s, the effect of the temporal location of the unfavorable information will be stronger than for low-D *S*s.

RESULTS

The mean ratings of each tape arranged by the interaction of Temporal Placement × Content Area Order are presented in Table 1. An analysis of variance (Winer, 1962) was performed on these data and is presented in Table 2. The temporal placement main effect was significant at the .01 level. Although the *t* value of .223 resulting from comparing the means of the first and second temporal position was not statistically significant, the order of magnitude of the three was consistent with the hypothesis. Neither the content area order nor the interaction of Placement × Order was statistically significant. Therefore, the ratings were rearranged into Table 3 which presents the interaction of Temporal Placement × Content Area. A second analysis of variance, presented in Table 4, was performed using temporal placement and content area as main effects. Again, the temporal placement was the only significant effect.

TABLE 2

SUMMARY OF ANALYSES OF VARIANCE: TEMPORAL PLACEMENT X CONTENT AREA ORDER

Source of variance	*SS*	*df*	*MS*	*F*
Temporal placement (A)	14.11	2	7.50	6.66*
Content area order (B)	7.58	5	1.52	1.43
A × B	6.27	10	.63	.59
Within cell	114.00	108	1.06	
Total	130.96	125		

* $p < .01$.

Since the content area and order were not significant, they were not used in the next analysis. The temporal placement and D-scale data were analyzed using multiple regression as a general variance accounting procedure (Cohen, 1968). A variation of Cohen's procedure was used to dummy variable code the temporal placement information and to put in the D scale as an interaction effect. The D scale was also to have been added in as a linear effect.

The first multiple regression used the dummy variable coding of the temporal placement as the independent variable and the rating as the dependent variable. The *F* value, from Table 5, was 4.08 which is significant at the .01 level. The multiple-correlation coefficient was .25 which meant that the temporal placement accounted for 6.25% of the variance in the ratings of the unfavorable information. The multiple regression which was to have added the D scale as a linear affect proved unnecessary. The correlation between

TABLE 3

MEAN RATING OF TAPED INTERVIEWS ARRANGED BY TEMPORAL PLACEMENT OF UNFAVORABLE INFORMATION × CONTENT AREA

Temporal placement	Content area									Row $\bar{X}$
	Academic			Personal-Social			Work history			
First third	Tape 5 3.57	Tape 6 3.14	3.35	Tape 11 3.43	Tape 12 3.43	3.43	Tape 17 3.14	Tape 18 3.00	3.07	3.28[a]
Second third	Tape 9 3.57	Tape 14 3.29	3.43	Tape 3 3.29	Tape 15 3.29	3.29	Tape 2 2.86	Tape 8 4.14	3.50	3.40[a]
Final third	Tape 10 3.86	Tape 13 4.14	4.00	Tape 4 3.43	Tape 16 4.29	3.57	Tape 1 4.43	Tape 7 4.14	4.29	4.05
Column $\bar{X}$			3.59			3.52			3.62	3.58

Note.—Rating scale of 1–7.

[a] $t = .223$ for comparison between mean for first third and for second third.

the rating and the D scale was −.01. Therefore, it could not have contributed as linear effect in a multiple-regression analysis using dummy variables.

In the second multiple regression, the independent variables were the temporal placement of the unfavorable information and the D scale of the rater coded as an interaction effect. The ratings again constituted the dependent variable. The results are reported in Table 6. Since only the first multiple regression produced results significantly different from zero, the analysis was terminated.

DISCUSSION

The major hypothesis in this study was that the earlier the unfavorable information occurred in the interview, the more unfavorable the ratings would be. This hypothesis was generally verified with the exception of Proposition *b*. With the unfavorable information occurring in the second third of the interview, the ratings were higher than when the unfavorable information came during the first, however, only slightly higher and clearly not significantly higher.

According to Hypothesis 2, the ratings should be the lowest (most favorable) when the unfavorable information was imbedded in the social-personal area and came during the first third of the interview. This hypothesis was not supported.

Based on Hypothesis 3, a negative correlation between the D-scale score of the raters and their ratings was expected. This hypothesis was not supported.

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE TEMPORAL PLACEMENT × CONTENT AREA

Source of variance	*SS*	*df*	*MS*	*F*
Temporal placement (A)	14.11	2	7.06	6.72*
Content area (B)	.93	2	.42	.40
A × B	1.60	4	.40	.38
Within cell	124.00	117	1.05	
Total	141.00	125		

* $p < .01$.

TABLE 5

SUMMARY OF ANALYSIS OF VARIANCE FOR MULTIPLE LINEAR REGRESSION WITH TEMPORAL PLACEMENT CODES AS INDEPENDENT VARIABLES AND APPLICANT RATINGS AS THE DEPENDENT VARIABLES

Source of variation	*SS*	*df*	*MS*	*F*
Due to regression	8.87	2	4.44	4.08*
Deviation about regression	113.83	123	1.09	
Total	142.70	125		

* $p < .01$.

TABLE 6

SUMMARY OF ANALYSIS OF VARIANCE FOR MULTIPLE LINEAR REGRESSION WITH TEMPORAL PLACEMENT AND D SCALE AS AN INTERACTION EFFECT AND APPLICANT RATINGS AS THE DEPENDENT VARIABLE

Source of variation	*SS*	*df*	*MS*	*F*
Due to regression	9.66	5	1.93	1.74
Deviation about regression	133.05	120	1.11	
Total	142.71	125		

Note.—The independent variables were temporal placement and the D scale as an interaction effect.

According to Hypothesis 4, the effect of temporal location of the unfavorable information would be stronger for high-D *S*s. This hypothesis also was not supported.

The primary conclusions for this study would appear to be: (*a*) Temporal placement of the unfavorable information is an important effect, confirming the conclusion of Bolster and Springbett (1961). (*b*) Temporal placement accounts for only a small percentage of the variance in the ratings. (*c*) The content area of the unfavorable information does not account for a significant amount of the variance, at least with the specific items of unfavorable information used in this study. (*d*) Content area order also does not account for a significant amount of the variance. (*e*) The influence of dogmatism of the rater as indicated by the D scale is called into question.

The findings of this study support Springbett's (1958) conclusion that if early information is not all favorable then the final decision is more likely to be negative. However, the finding that when the unfavorable information comes during the middle portion of the interview the ratings are not significantly different from those obtained when it occurs during the first portion, suggests a need for further study of the applicability of Bruner's (1957) stages in a decision sequence to the selection interview.

The findings of this study are in agreement, as well, with Bolster and Springbett's (1961) conclusion that there was a modified primacy effect, even though their study involved ratings based on written information. If the primacy effect operated in the usual manner, and in accordance with Bruner's (1957) stages, all three propositions of Hypothesis 1 should have been verified. It would be interesting to know if this is an effect of having used taped interviews instead of written information. It might also be a function of not having used intermediate ratings as Crowell's (1961) study suggests.

The rejection of Hypothesis 3 is not in keeping with Wheeler's (1969) findings of a negative correlation between the D-scale score of the raters and their ratings of specific pieces of interview data. This could have resulted from the fact that Wheeler's (1969) *S*s were older and more experienced than those of the present study. There is a need for additional research to unravel the relationship between these variables.

The time required to develop tapes such as those used in this study is one of the major limitations of the use of audio tapes. However, the use of recordings of simulated selection interviews provides a way to control many variables, such as content, voices, and so on, that cannot be controlled in a real interview.

REFERENCES

BOLSTER, B. I., & SPRINGBETT, B. M. The reaction of interviewers to favorable and unfavorable information. *Journal of Applied Psychology*, 1961, **45**, 97–103.

BRUNER, J. S. On perceptual readiness. *Psychological Review*, 1957, **64**, 123–152.

COHEN, J. Multiple regression as a general data-analytic system. *Psychological Bulletin*, 1968, **70**, 426–443.

CROWELL, A. H. Decision sequences in perception. Unpublished doctoral dissertation, McGill University, 1961. Cited by E. C. Webster, *Decision-making in the employment interview*. Montreal: Eagle, 1964.

ROKEACH, M. *The open and closed mind*. New York: Basic Books, 1960.

SPRINGBETT, B. M. Factors affecting the final decision in the employment interview. *Canadian Journal of Psychology*, 1958, **12**, 13–22.

WEBSTER, E. C. *Decision-making in the employment interview*. Montreal: Eagle, 1964.

WHEELER, R. W. A study of the relationship between selected interviewer variables and the interpretation of interview information. Unpublished doctoral dissertation, University of Houston, 1969.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received December 31, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 143–150

STABILITY OF VOCATIONAL INTERESTS WITHIN NONPROFESSIONAL OCCUPATIONS[1]

ALAN W. LAU[2] AND NORMAN M. ABRAHAMS

Naval Personnel and Training Research Laboratory, San Diego

The use of empirically derived occupational interest scales assumes that individuals' scores and criterion group characteristics remain stable over relatively long periods of time. While these assumptions have been tested with inventories used for professional occupations, they have not been tested in the nonprofessional interest domain. This study compared Navy Vocational Interest Inventory scores received by Navy enlisted men in a variety of occupational specialties with their retest scores obtained 5 yr. later. Results show substantial reliability of individual scores, paralleling reliabilities obtained with the Strong Vocational Interest Blank. Comparisons of interest profiles of criterion groups tested in 1951 with those of men entering the same specialties 13 yr. later also showed considerable stability. Findings should be generalizable to the Minnesota Vocational Interest Inventory, a revision of the original inventory.

In comparison to the measurement of interests among men in professional occupations, men in nonprofessional occupations have been virtually neglected. Clark (1961) reported a pioneering effort in this area when he developed an interest inventory for Navy recruit classification and assignment. This inventory, the Navy Vocational Interest Inventory (NVII), emphasized activities which differentiated between men in various skilled trades and technical occupations; less emphasis was placed on professional or unskilled activities. Later, Clark and Campbell (1965) published an abbreviated version of the earlier inventory, the Minnesota Vocational Interest Inventory (MVII), for use in vocational guidance among nonprofessional occupations.[3]

Items from the NVII which differentiated between the interests of a general reference group (Navy men in general) and criterion groups of men who were members of specific Navy occupational specialties were selecte for scoring keys. Using this rationale, Clar constructed a total of 19 occupational keys In addition to these empirical occupationa keys, Clark developed homogeneous (or area scales. The area scales draw together item with a common core into homogeneous groupings, and provide relatively pure measures o interest traits or factors.

The use of empirically derived occupationa scales requires two assumptions of stability First, since scales are constructed on individuals tested as members of occupationa groups, it is assumed that they would hav responded essentially the same had they bee tested prior to occupational choice. In othe words, it is assumed that individuals' score remain stable over time. The second assumption is that men entering a given occupation several years after scale development have the same interests as the original criterion group. For instance, if mechanics tested several years ago differed considerably from mechanics tested today, scales constructed on the earlier group would be inappropriate.

These assumptions have been adequately tested with inventories used for professional occupations. However, with nonprofessional occupations these assumptions have not been examined. The only reported study on reliability of the NVII scales, though rather limited by the short retest interval of 30 days, indicated fairly high reliability for a sample

---

[1] The opinions expressed are those of the authors and do not necessarily reflect those of the Navy Department.

[2] Requests for reprints should be sent to Alan W. Lau, Research Psychologist, Personnel Measurement Research Department, Naval Personnel and Training Research Laboratory, Department of the Navy, San Diego, California 92152.

[3] In developing the Minnesota inventory, some editing was done to simplify the wording of the original items, and 32 triads were eliminated. The items included in the area scales of the two inventories are identical.

TABLE 1

NUMBER OF MEN BY INITIAL SCHOOL ASSIGNMENT WHO RETURNED NVII

Rating	Sent NVII	Returned NVII
Hospitalman	26	20
Electronics Technician	22	20
Yeoman	20	18
Storekeeper	12	8
Engineman	19	18
Machinist's Mate	31	27
Submarine[a]	125	97
Total	255	208

Note.—NVII = Navy Vocational Interest Inventory.
[a] Of the 97 men in the Submarine school, 62 were in mechanical occupational specialties; 31 were in electrical-electronics specialties; 4 men were in other miscellaneous specialties.

of 98 trade school students (Clark, Gee, & Perry, 1953).

The bulk of research with the inventory has been toward establishing the ability of scales to differentiate men in different occupations. Other relevant research has indicated that NVII test scores are related to Navy technical (Class A) school achievement (Clark et al., 1953; Spies, 1966), job performance of IBM workers (Ghei, 1960), and job satisfaction of Navy Yeomen (Perry, 1960). Recent research conducted by Lau (1969) has shown that when Navy enlistees are engaged in vocational activities which are consonant with their interests, they are more likely to reenlist and to receive higher performance ratings and technical school grades.

While a good deal of research has been directed at validity, the two assumptions of stability have been virtually ignored. Consequently, the present study sought to furnish evidence on the stability of individual scores and of occupational groups on the scales of this inventory. This study examines the 5-yr. test-retest reliability of the occupational and area scores of the NVII and the stability of occupational scale profiles for men entering occupational groups 13 yr. apart. If it can be shown that scores are reasonably stable and the original scales are still appropriate, the inventory may be used with more confidence in vocational counseling.

METHOD

Subjects

In 1964, the NVII was administered to 1,009 recruits at seven technical Navy schools prior to the beginning of training. These seven schools were Hospitalman, Electronics Technician, Yeoman, Storekeeper, Engineman, Machinist's Mate, and Submarine. The interest test was administered on an experimental basis and scores were not used for the assignment of men to the various schools.

Procedure

By searching Navy personnel records in 1969, 255 of the original testees were identified either as reenlistees or as men who had extended their first enlistments. These men were followed up, and NVII booklets were mailed to them at their duty stations. Completed inventories were received from 208 men (82%). Table 1 presents the number of men by initial school assignment who returned the inventory.

TABLE 2

MEANS AND STANDARD DEVIATIONS ON RELEVANT OCCUPATIONAL KEYS FOR CLARK'S CRITERION GROUPS AND THE 1964 OCCUPATIONAL GROUPS

Key	No. of items	Original criterion groups			1964 Occupational groups		
		N	$\bar{X}$	*SD*	*N*	$\bar{X}$	*SD*
Hospitalman	69	239	70.9	13.9	266	77.4*	8.9
Electronics Technician	71	217	67.1	8.4	116	64.2*	8.3
Engineman	70	237	59.1	6.6	113	58.4	6.5
Machinist's Mate	61	254	60.8	9.4	88	51.7*	8.3
Storekeeper	64	146	65.3	10.3	88	66.7	9.2
Yeoman	70	227	67.1	11.0	99	71.9*	8.1

Note.—Raw scores have been transformed to standard scores ($\bar{X} = 50$, $SD = 10$) based on the standardization sample of Navy men in general.
* The difference between mean scores of the original and 1964 groups is significant at $p = .01$.

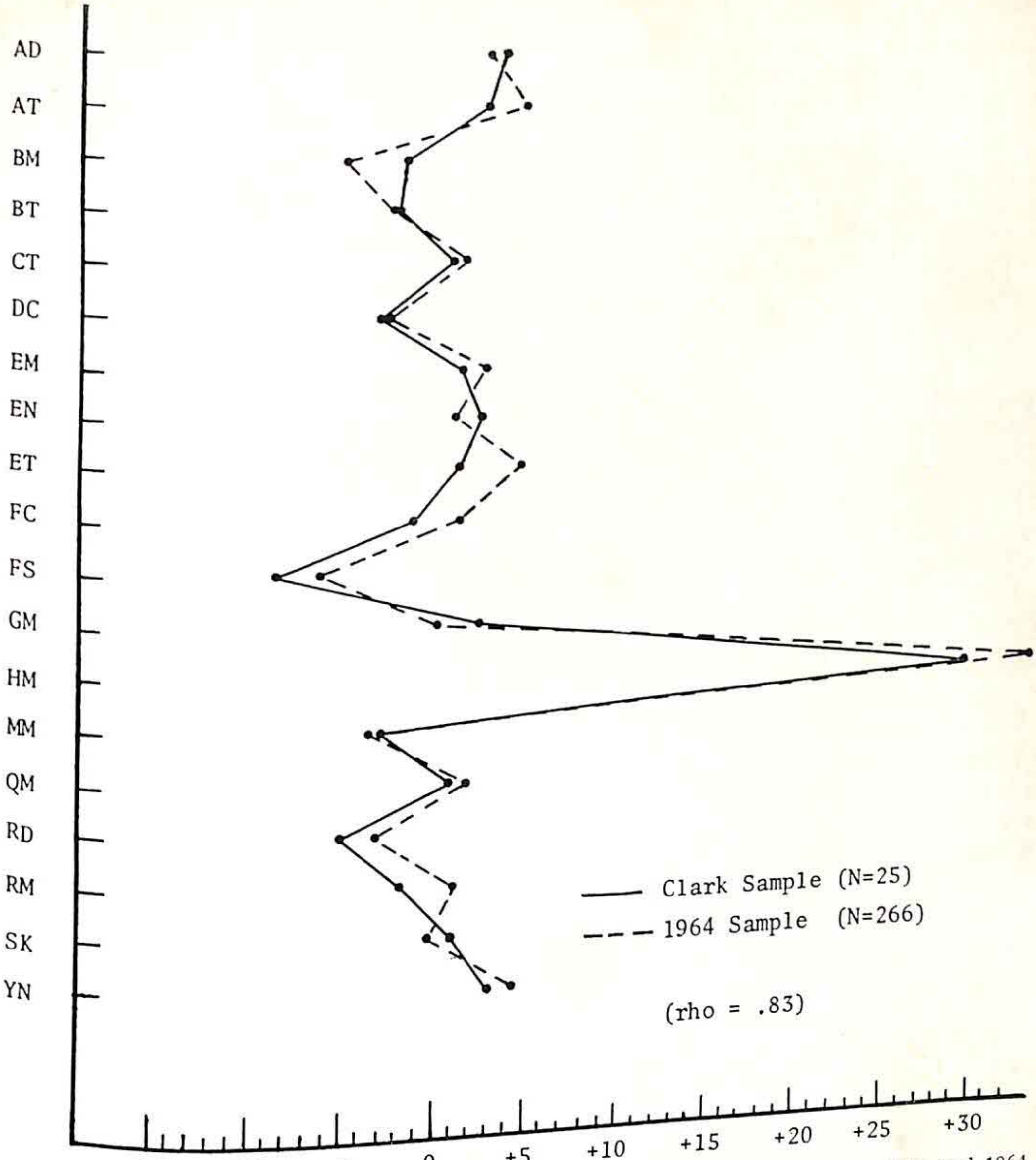


FIG. 1. Mean profile on Navy Vocational Interest Inventory occupational scales for 1951 and 1964 Hospitalmen. (Table entries are in raw score units.)

NVII scores were obtained on nine area scales and 19 occupational keys. From Clark's original work, a relevant key (i.e., a key developed specifically for men in that occupational specialty) was available for all but the Submarine school.

RESULTS AND DISCUSSION

Thirteen-Year Occupational Group Stability

The first analysis addressed the question of whether the interests of men in Navy occupational groups have changed since Clark's original data were collected. If the interests of occupational groups have substantially changed, the original occupational keys would not be of much value for differentiating men currently in these occupational specialties.

To examine this question, mean scores on occupational keys obtained by men in the occupational specialties tested in 1964 were compared with scores obtained by men in the same specialties in 1951.[4] For this analysis, mean profiles for small randomly selected samples of men in the six specialties were extracted from Clark et al. (1953) and compared with the mean profile for men in the same specialty in 1964.

---

[4] Since Clark had not constructed scales for men in the Submarine school, these men could not be included in this analysis. Area scale scores obtained by Clark's original criterion groups are also not available.

The 1951 and 1964 profiles revealed a great deal of occupational group stability within each of the six occupational specialties. Figure 1, which shows the profile for Hospitalman samples tested 13 yr. apart, is presented as an illustration. To measure the degree of profile similarity, Spearman rank-difference correlations for each of the six Navy occupational groups were computed. Profile correlations were as follows: Hospitalman ($\rho = .83$), Electronics Technician ($\rho = .93$), Engineman ($\rho = .95$), Machinist's Mate ($\rho = .91$), Storekeeper ($\rho = .91$), and Yeoman ($\rho = .61$).

Scores earned on all occupational scales by the total sample of men tested in each of the original criterion groups were not available. However, Clark et al. (1953) does present each criterion group's score on its own scale. For a further stability assessment, mean scores on their own scales (i.e., the scale constructed on this group's responses) for Clark's criterion group samples are compared with those of men in the 1964 occupational specialties in Table 2.

While four of the six comparisons yield significant differences, only two show the direction of change that might cause concern, that

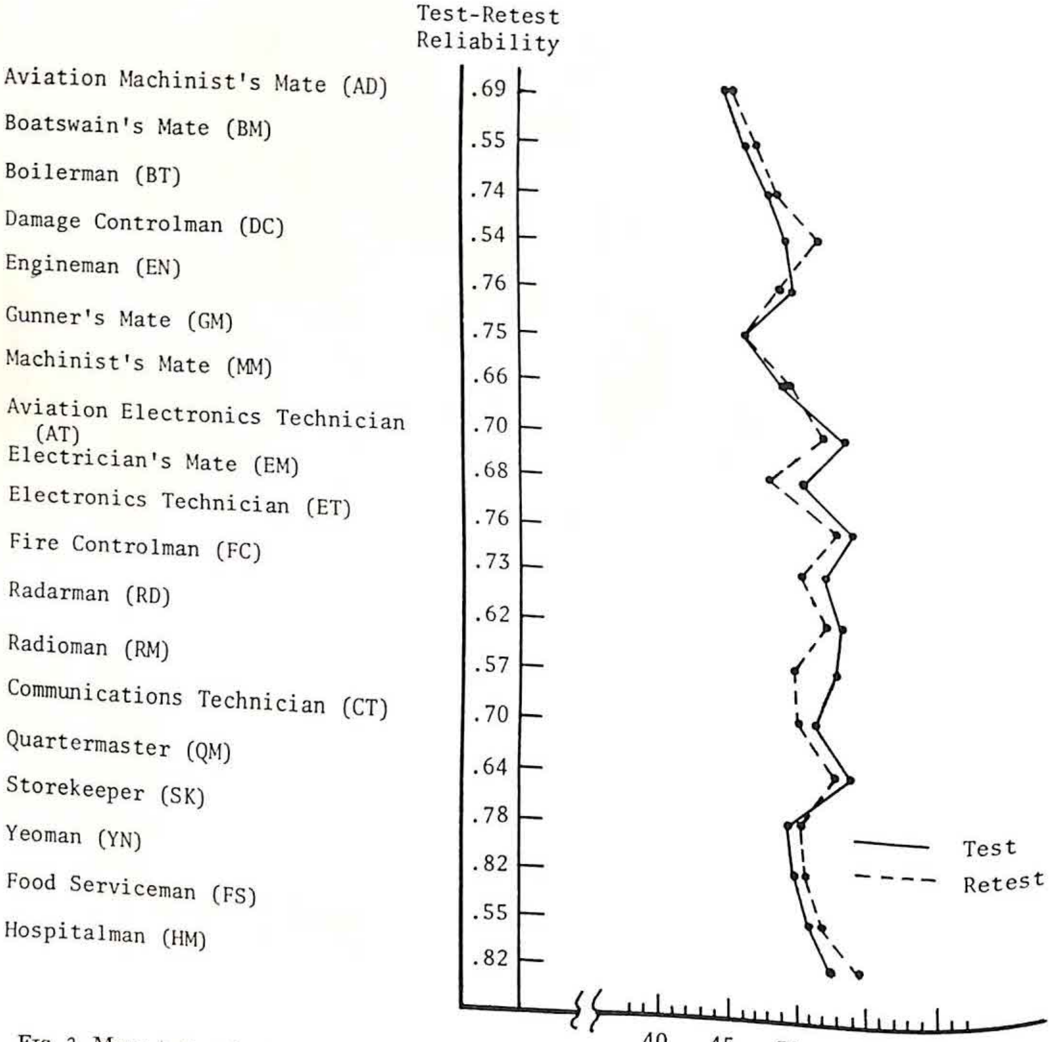


FIG. 2. Mean test and retest profile of Navy Vocational Interest Inventory occupational keys for all Navy occupational samples. ($N = 208$; raw scores have been transformed to standard scores [$X = 50$, $SD = 10$] based on the standardization sample of men in general.)

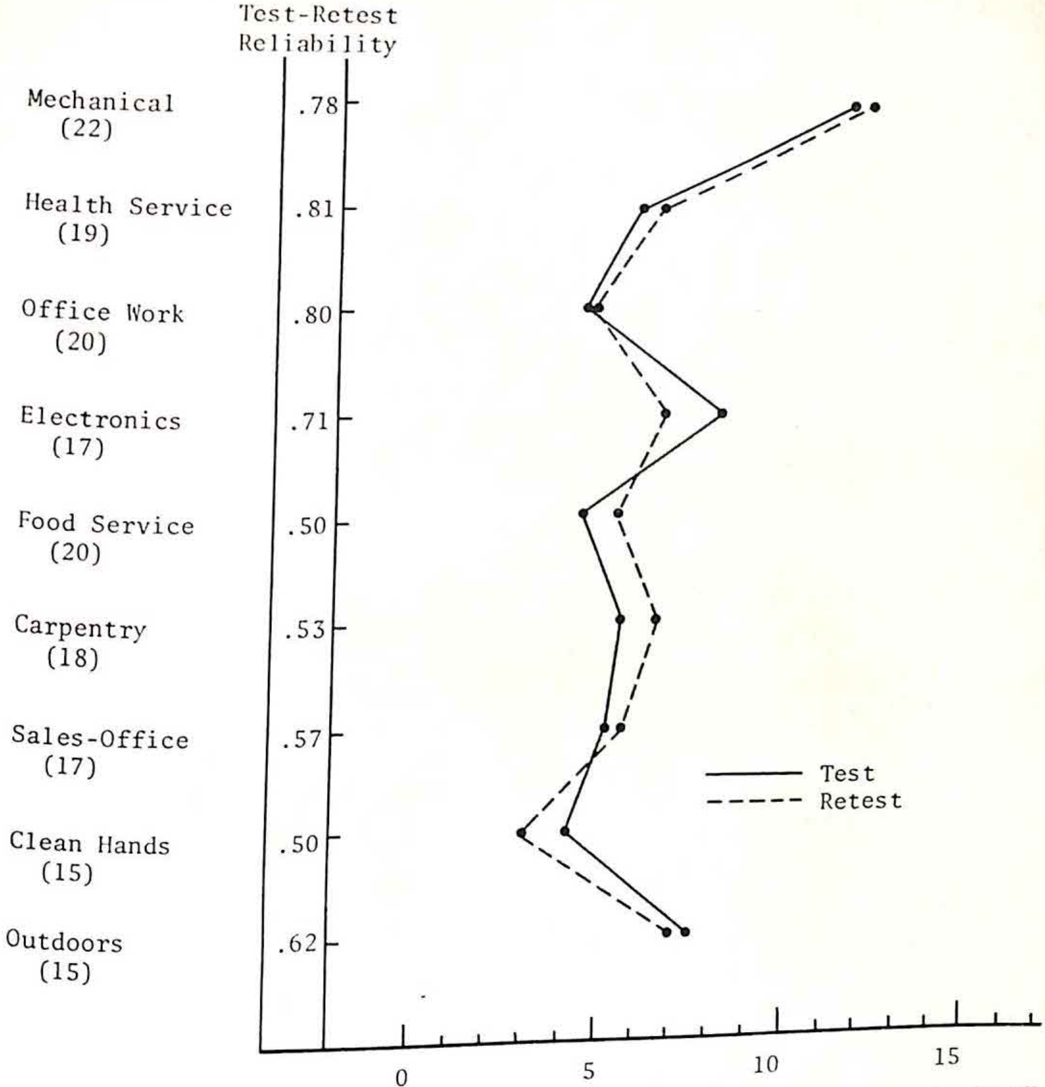


FIG. 3. Mean test and retest profile of Navy Vocational Interest Inventory area scales for all Navy occupational samples. ($N = 208$; table entries are in raw score units; number of items in each area scale is given in parentheses.)

is, decreases. For Electronic Technicians the decrease from the original criterion group is less than one-half of a standard deviation and for Machinist's Mates about one standard deviation. These comparisons, however, may be considered rather demanding, since they isolate one scale from a large set of scales and give no indication of overall profile similarity. For example, the comparison presented earlier for an original sample of Machinist's Mates with the new sample yielded a rank-order correlation of .91 between profiles.

A further analysis may be made by examining the ability of each relevant scale to differentiate the 1951 and 1964 occupational groups from the original reference group of Navy men in general. The percentage overlap between the distributions of the reference group and each of the original and new occupational groups on their own scale was computed. For the 1951 groups the average percentage overlap with the reference group is 46, compared to 48 on the new groups.

These results taken together strongly suggest that the original occupational keys are still appropriate for measuring the interests of men in the same occupations. However, before this conclusion can be made, stability of indi-

TABLE 3

NVII TEST-RETEST SCORES AND SCALE RELIABILITIES FOR OCCUPATIONAL GROUPS AND TOTAL SAMPLE

Scale	Occupational group																								
	Mechanical[a]					Electrical-Electronics[b]					Clerical[c]					Health service[d]					Total sample[e]				
	Test			Retest		Test			Retest		Test			Retest		Test			Retest		Test			Retest	
	ru	$\bar{X}$	*SD*	$\bar{X}$	*SD*	*ru*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	*ru*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	*ru*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	*ru*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
Occupational																									
Aviation Machinist's Mate	.50	50.7	6.7	50.1	8.4	.50	43.7	6.5	43.8	6.7	.67	35.7	6.5	38.3	7.4	.59	36.1	4.4	38.2	5.9	.69	45.6	8.8	45.8	9.0
Aviation Electronics Technician	.60	51.5	8.9	50.4	8.6	.41	65.3	6.1	62.4	6.8	.09	45.2	5.8	44.6	5.4	.02	52.9	5.7	50.3	5.9	.70	54.0	10.1	52.5	9.6
Boatswain's Mate	.52	48.7	9.2	49.6	8.8	.51	40.8	7.3	42.7	6.9	.06	49.6	6.8	49.3	7.0	.65	49.3	7.7	48.9	9.5	.55	47.0	9.0	47.8	8.7
Boilerman	.67	54.5	9.0	53.8	9.3	.54	45.1	7.9	45.4	7.2	.59	41.1	6.4	41.7	9.6	.56	37.8	4.7	40.6	7.2	.74	48.8	10.2	48.9	10.2
Communications Technician	.54	46.6	8.0	45.5	8.9	.55	55.4	7.1	54.3	7.5	.47	61.9	6.1	59.0	8.9	.67	58.6	6.6	57.2	6.9	.70	51.9	9.5	50.5	10.0
Damage Controlman	.49	52.2	7.7	54.4	7.8	.62	44.7	5.9	48.0	6.3	.15	49.4	5.8	53.2	6.9	.38	49.4	7.6	51.3	7.7	.54	49.7	7.7	52.4	7.8
Electrician's Mate	.50	51.0	8.2	48.3	8.6	.55	60.8	5.3	56.5	8.7	.19	39.6	4.9	39.6	5.6	.49	42.3	8.1	42.0	5.7	.68	51.0	10.0	48.5	9.7
Engineman	.57	*55.7*	7.9	*54.6*	9.1	.55	49.2	7.3	48.6	6.8	.66	36.1	5.4	37.7	8.3	.60	38.1	3.9	39.8	7.1	.76	49.8	10.4	49.4	10.5
Electronics Technician	.68	51.8	10.5	50.8	10.0	.46	*67.2*	5.8	*64.2*	7.0	.43	44.1	5.8	44.6	5.9	.23	55.0	5.8	51.2	7.1	.76	54.7	11.4	53.2	10.8
Fire Controlman	.65	50.8	9.1	49.0	9.1	.51	64.6	6.2	60.8	9.2	.28	44.2	5.0	44.4	6.0	.32	45.6	5.7	44.3	5.0	.73	52.7	10.5	50.8	10.3
Food Serviceman	.59	50.2	8.0	51.1	9.0	.43	48.7	7.6	51.1	9.1	.68	54.2	8.7	56.4	11.1	.36	57.2	10.6	53.3	6.9	.55	51.1	8.6	52.0	9.3
Gunner's Mate	.69	52.1	9.2	50.8	10.2	.76	44.7	8.1	44.8	7.7	.49	39.5	5.3	42.1	7.0	.74	35.1	4.8	38.5	7.5	.75	47.0	10.1	47.0	10.0
Hospitalman	.72	50.0	9.7	52.6	10.4	.64	51.8	10.2	53.3	8.7	.64	46.4	5.6	48.4	7.5	.81	*79.8*	7.1	*75.9*	10.3	.82	52.7	12.7	54.4	12.0
Machinist's Mate	.53	*54.7*	8.3	*54.8*	8.1	.52	47.9	7.4	47.3	6.1	.29	39.2	6.0	41.3	7.8	.44	43.6	5.7	45.1	6.7	.66	50.0	9.5	50.2	9.0
Quartermaster	.64	50.9	8.5	50.5	9.2	.54	56.3	8.1	54.5	7.5	.53	58.7	6.8	57.0	6.6	.41	60.5	5.9	57.2	7.9	.64	54.2	8.7	53.0	8.8
Radarman	.53	51.5	8.6	51.2	8.2	.31	64.3	6.3	60.4	8.1	.42	49.4	7.5	50.4	6.4	.20	45.7	7.3	44.7	6.2	.62	53.7	9.9	52.7	9.1
Radioman	.43	49.9	8.8	46.4	8.1	.44	62.3	6.2	57.8	8.8	.24	51.1	5.7	51.1	6.8	.48	51.6	7.4	49.9	5.3	.57	53.2	9.2	50.1	9.1
Storekeeper	.41	45.3	6.8	46.2	7.3	.70	46.3	5.8	46.8	6.4	.76	*69.8*	6.9	*66.0*	10.2	.69	55.6	8.0	57.5	9.1	.78	49.8	10.8	50.1	10.5
Yeoman	.52	45.3	6.6	46.3	7.5	.73	47.4	6.5	47.4	6.6	.48	*73.0*	6.4	*69.0*	9.7	.71	56.9	9.0	58.2	8.1	.82	50.7	11.6	50.8	11.0
Area																									
Mechanical	.59	*15.1*	5.1	*14.7*	5.6	.67	12.7	5.6	13.0	5.4	.48	2.3	1.6	4.6	4.7	.84	3.2	3.1	4.2	4.6	.78	11.6	6.9	11.9	6.8
Health Service	.71	4.8	4.2	5.7	4.4	.72	6.0	4.6	6.4	3.8	.58	2.5	2.8	3.2	3.2	.83	*17.0*	3.1	*14.9*	4.6	.81	5.9	5.5	6.4	5.0
Office Work	.48	2.4	3.0	2.8	3.2	.71	3.3	3.0	3.3	3.3	.76	*15.0*	4.3	*12.8*	5.7	.62	4.8	3.7	6.8	4.4	.80	4.5	5.3	4.6	5.1
Electronics	.43	8.4	3.7	6.7	3.5	.71	*12.6*	3.1	*10.7*	4.1	.08	2.7	1.8	3.4	2.5	.74	2.7	3.1	2.2	2.1	.71	8.1	4.8	6.8	4.4
Food Service	.51	4.3	3.2	5.2	3.6	.51	3.9	3.2	5.4	3.1	.55	4.8	3.6	5.6	4.6	.55	6.2	4.3	5.3	2.9	.50	4.4	3.4	5.3	3.6
Carpentry	.47	6.2	3.5	7.7	3.5	.19	3.1	2.0	4.4	2.6	.44	6.9	2.7	8.4	2.3	.39	4.1	2.8	5.9	3.4	.53	5.3	3.4	6.8	3.4
Sales-Office	.52	4.5	2.3	4.9	2.5	.57	5.6	2.8	6.1	2.6	.60	5.6	2.2	6.5	2.5	.19	7.4	1.9	7.4	2.5	.57	5.2	2.5	5.7	2.6
Clean Hands	.38	3.3	2.0	2.3	1.9	.41	3.9	2.1	3.0	2.0	.40	7.0	1.7	5.0	2.0	.20	4.6	2.2	3.9	2.5	.50	4.1	2.4	3.0	2.2
Outdoors	.54	8.8	2.6	8.3	2.8	.53	7.2	2.3	6.6	2.2	.27	4.9	2.2	5.5	2.3	.71	5.5	3.2	5.8	2.7	.62	7.6	2.9	7.3	2.8

Note.—Raw scores have been transformed to standard scores ($\bar{X} = 50$, $SD = 10$) based on the standardization sample of Navy men in general, and means for relevant NVII keys are italicized. NVII = Navy Vocational Interest Inventory.

[a] N = 109.

[b] N = 52.

[c] N = 27.

[d] N = 20.

[e] N = 208.

vidual scores over time on the same individuals must be examined through test-retest correlations.

Five-Year Scale Reliability

The next two analyses sought to assess the reliability of scale scores over a 5-yr. interval. The first analysis compared test and retest scores for the entire sample, while the second analysis examined scale reliability within occupational specialties.

Figure 2 presents a test and retest mean profile and scale reliabilities on the occupational scales for the entire sample. Figure 3 presents parallel information on the area scales. Further, the last column of Table 3 provides test-retest means, standard deviations, and reliabilities for the entire sample. Clearly, there is considerable stability for both occupational and area scales. Occupational scale test-retest reliabilities range from .54 to .82 with a median value of .70. Area scales, which generally contain about one-third the number of items in the occupational scales, range in reliability from .50 to .81 with a median of .62. The median 30-day reliabilities reported by Clark et al. (1953) for occupational and area scales are .82 and .83, respectively. While the 5-yr. reliabilities are, as would be expected, considerably lower than the 30-day values, the median occupational scale reliability is comparable to the median value of .68 obtained with the SVIB over a 3-yr. interval on 189 Harvard freshmen (Campbell, 1966). To insure the generality of the NVII test-retest results, a parallel follow-up of men who left the Navy is in progress.

The final analysis examined the scale reliabilities within occupational specialties. For this comparison, the constructed occupational groups were mechanical ($N = 109$), electrical-electronics ($N = 52$), clerical ($N = 27$), and health service ($N = 20$). The results of this analysis, including means, standard deviations, and test-retest correlations for all scales and groups are given in Table 3.

These data indicate several interesting findings. First, the test-retest correlations within occupational groups are generally lower than for the total sample. Inspection of the standard deviations reveals, as would be expected, lower variabilities within occupational groups than for the total sample, thereby restricting the magnitude of test-retest correlations. Second, inspection of test and retest mean scores on the relevant area and occupational scales within each occupational group shows a consistent and significant decrease at the time of retesting for all but the mechanical group. Concomitantly, on over half of the relevant scales, variability significantly increases after 5 yr. in the occupational specialty, suggesting that actual experience in a specialty expands the range of interest in that work.

A reasonable hypothesis for such increased variability and decreasing means on relevant scale scores seems to lie in the trial and error process of ultimate vocational choice. A group entering a given occupation presumably would be subjected to common occupational stereotypes, resulting in their being somewhat homogeneous in their attraction to that field. Actual exposure to the work itself provides a more meaningful basis for determining likes and dislikes. With experience in place of stereotypes, it would be anticipated that some men would be disappointed and disenchanted while others would be satisfied; thus, increased variability and decreased magnitude of scores on that occupational scale would be expected upon retesting.

CONCLUSIONS

With respect to the stability of nonprofessional interest scales and groups, the following conclusions seem warranted:

1. The reliability over extended time intervals of nonprofessional occupational interest scales parallels that obtained with professionally oriented measures and appears to be high enough to justify their use in counseling.

2. The interest profiles of men who entered the same nonprofessional occupations 13 yr. after the initial scale development were quite similar to those of the original criterion groups. It appears that these scales, constructed several years ago, are still appropriate for groups tested more recently.

3. Actual exposure to the work of an occupational specialty appears to lower, slightly, the general level of interest and increase the range of scale scores.

REFERENCES

Campbell, D. P. *Manual for the Strong Vocational Interest Blanks for men and women.* Stanford: Stanford University Press, 1966.

Clark, K. E. *The vocational interests of nonprofessional men.* Minneapolis: University of Minnesota Press, 1961.

Clark, K. E., & Campbell, D. P. *Manual for the Minnesota Vocational Interest Inventory.* New York: Psychological Corporation, 1965.

Clark, K. E., Gee, H. H., & Perry, D. *Measurement of interest patterns.* (Project N6ori-212, Task Order III, Annual Report) Minneapolis: University of Minnesota, 1953.

Ghei, S. Vocational interests, achievement, and satisfaction. *Journal of Counseling Psychology,* 1960, 7, 132–136.

Lau, A. W. The Navy Vocational Interest Inventory as a predictor of school performance, job performance, and reenlistment. Paper presented at the 11th Annual Conference of the Military Testing Association, New York, September 1969.

Perry, D. Problems of item form and criterion group definition. In W. L. Layton (Ed.), *The Strong Vocational Interest Blank research and uses.* Minneapolis: University of Minnesota Press, 1960.

Spies, C. J. *Some non-intellectual predictors of classroom success.* (Tech. Rep. No. 10, Contract Nonr-816[14]) St. Louis: Washington University, October 1966.

(Received January 15, 1970)

(Continued from page 109)

Personal Orientation Inventory as a Predictor of Success in Peace Corps Training: Michael J. Uhes, and John Shybut*: Department of Psychology, Northern Illinois University, DeKalb, Illinois 60115.

Influence of Expectation of Task Duration on Efficiency of Muscular Activity: Stjepan Vidacek, and Julius Wishner*: Department of Psychology, University of Pennsylvania, 3813-15 Walnut Street, Philadelphia, Pennsylvania 19104.

Face to Face Interaction in the Peer Nomination Process: Arie Y. Lewin*, Peter Dubno, and William G. Akula: Graduate School of Business Administration, New York University, 100 Trinity Place, New York, New York 10006.

Factors Related to Acceptance and Refusal of Employment Among Temporary Help Applicants: Martin J. Gannon*, and Uri Brainin: Behavioral Science Division; Department of Business Administration, University of Maryland, College Park, Maryland 20743.

Sound and Noise "Pollution" and Health: Community Parameters: Paul Cameron*, Donald Robertson, and Jeffry Zaks: Department of Psychology, University of Louisville, Louisville, Kentucky 40208.

Relationship of Initiating Structure and Job Performance as Moderated by Consideration: Robert C. Cummins*: Department of Communication and Organizational Behavior, General Motors Institute, 1700 West Third Avenue, Flint, Michigan 48502.

Examination of the Relationship between the Accuracy and Favorability of Ratings: Michael E. Gordon*: Department of Industrial and Personnel Management, College of Business Administration, University of Tennessee, Knoxville, Tennessee 37916.

Informational Social Influence and Product Evaluation: Joel B. Cohen*, and Ellen Golden: Department of Business Administration, University of Illinois, Urbana, Illinois 61801.

Importance of Contrast Effects in Employment Interviews: Kenneth N. Wexley*, Gary A. Yukl, Steven Z. Kovacs, and Raymond E. Sanders: Department of Psychology, University of Akron, Akron, Ohio 44304.

Occupational Preferences: Perceptions of Valence and Instrumentality, and Objective Data: John P. Wanous*: Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.

---

* Asterisk indicates author for whom address is supplied.

Journal of Applied Psychology
1971, Vol. 55, No. 2, 151–156

ASSIMILATION OF NEW MEMBERS IN A LARGE AND A SMALL CHURCH[1]

ALLAN W. WICKER[2] AND ANNE MEHLER

University of Wisconsin, Milwaukee

Consistent with behavior setting theory, new members of a small church (338 members, $n = 26$) reported greater assimilation than new members of a comparable large church (1,599 members, $n = 40$). The assimilation measure was a composite based on participation in church activities, felt obligation to participate, sense of belonging to the church, and acquaintance with church members and officers. College students asked to role play new members of either the small church ($n = 21$) or the large church ($n = 21$) did not differ in reported assimilation. The 2 × 2 analyses of variance (large vs. small church, new vs. established members) of self-report data on participation and church records of attendance and contributions revealed greater support for church activities by members of the small church and by established members.

The present article is one of a series of studies examining various aspects of Barker's (1960, 1968) behavior setting theory. According to the theory, the behaviors and experiences of members of voluntary organizations are influenced by the degree to which an organization's activities are undermanned. In an undermanned activity there are many jobs to be done relative to the available population, and thus if members are to obtain the satisfactions the activity provides, they must work harder, engage in a wider variety of tasks, and assume more responsibility than would be the case if there were an abundance of personnel. Greater responsibility for the activities leads the members to experience involvement, challenge, and self-worth. Also, members in undermanned activities are likely to recruit others, to lower requirements for admission, and to share their responsibilities.

Research on the theory has dealt with high school extracurricular activities such as school plays and athletic contests (Baird, 1969; Barker & Gump, 1964; Wicker, 1968, 1969a; Willems, 1967) and church activities such as worship services, business meetings, and social events (Wicker, 1969b). The research has shown that the activities of small schools and churches are undermanned relative to those of large schools and churches. Consistent with the theory, members of the small organizations tend to show more behavioral support for their organization's activities (e.g., they serve in more positions of responsibility, contribute more money, attend more frequently) and they tend to have more feelings of self-improvement, challenge, importance, and concern about activities.

One aspect of behavior setting theory which has not been previously studied is the effect of undermanning on the assimilation of new members into an organization's activities. The present study examines this relationship in a large and a small church. It would follow from the theory that the need for additional personnel in the activities of the small church would lead members to seek out potential members and to make them feel welcome and perhaps even obligated to participate. Also, the newcomers would be rapidly drawn into the relatively large number of positions of responsibility available to them. Increased participation would lead new members to meet and interact with many of the church's members and officers. And later, new members themselves might recruit others to participate. In contrast, in the large church which had a much larger pool of available personnel to carry out its activities, there would be less need for members to recruit new members, and the newcomers in this situation would feel less welcome and obligated to participate. There would be fewer positions of responsibility open to the new members, and they would be less likely to serve in them. Their acquaintance with other members and officers would be less intimate, and they would feel little pressure to recruit others.

---

[1] This research was supported by Public Health Service Grant No. 1 RO3 MH15798-01 and by a grant from the Graduate School of the University of Wisconsin, Milwaukee. Use of University of Wisconsin computing facilities was made possible through support from the National Science Foundation, other United States Government agencies, and the Wisconsin Alumni Research Foundation (WARF) through the University of Wisconsin Research Committee.

The assistance of the Reverends Earl Allen, Roger Becker, and Lee Smith, and the cooperation of the members of their congregations, is gratefuly acknowledged. Thanks are also due to J. Richard Hackman and Joseph McGrath for comments on an earlier version of this article, and to Marguerite Sowle for assistance in data collection. Portions of this article were presented at the Rocky Mountain Psychological Association Convention, Salt Lake City, Utah, May 1970.

[2] Requests for reprints should be sent to Allan W. Wicker, who is now at the Department of Psychology, University of Illinois, Champaign, Illinois 61820.

The present study also examines in a factorial design the relationship of church size and of recency of membership to members' support for church activities. A previous study of church members who were selected without regard to length of membership (Wicker, 1969b) provided a source of data on longer term or established members with which to compare the present data on new members. Thus, it was possible to test main effects of size and recency of membership, as well as their interaction. While Barker (1960) does not predict an interactive effect, it would be of interest to determine, for example, whether the difference between levels of support for church activities by new and old members is the same or different in large and small churches.

METHOD

Subjects

Data were obtained from members of a small United Methodist Church (338 members) and from a comparable large church of the same denomination (1,599 members). Both churches were located in metropolitan Milwaukee, Wisconsin. More complete descriptive data on the churches, and evidence that the activities of the small church are undermanned relative to those of the large church are presented in Wicker (1969b). In each church, membership records were used to identify all married couples who had joined the church between $3\frac{1}{2}$ and 30 mo. prior to the investigation. Persons who had participated in a study conducted a year earlier were not included. There were 30 people in the small church and 81 in the large church who met the criteria. Thirty persons were randomly eliminated from the large church, leaving 51 in the sample.

The Ss were first contacted by a personal letter from their church pastor; questionnaire data were obtained in group meetings at the churches or individually at Ss' homes. Questionnaires were completed by 26 Ss from the small church (87% of those contacted) and by 40 Ss in the large church (78%). The difference in participation of the members of the two churches does not approach significance ($\chi^2 = .89$). The average length of membership for the small church participants was 15.0 mo., compared to 17.8 mo. for the large church. The two samples were highly similar in average age, number of children, sex distribution, percentage who had attended another church in the 2 yr. prior to joining the present church, reported attendance at the previous church, responsibilities held in previous church, and satisfaction with previous church.

Questionnaire Material

1. New member questionnaire. Questions dealt with the following: demographic variables (age, sex); involvement in and satisfaction with the church that *S* previously attended (if any); rating of importance of various factors in choosing a church; beliefs about differences in large and small churches; and a series of questions to measure assimilation into the church. Six of the assimilation questions dealt with three activities common to both churches: church choir, social clubs for married couples, and the women's or men's organization. For each of the three activities, Ss were asked how frequently: (*a*) they had been invited to join or attend, (*b*) other members had offered to accompany them, (*c*) they had attended, and (*d*) they had asked someone else to attend. Also for each of the three activities, Ss indicated on 4-point scales how welcome they felt and the extent to which they felt obligated to participate. For the above questions, responses were summed across the three activities. An additional index of felt obligation was obtained from Ss' responses to questions about whether they felt they should help by teaching a children's church school class, working in other church activities, being an officer in a church organization, and making a monetary pledge to the church. The index was the number of "yes" responses to the four questions. Finally, a general question about assimilation was the following: "To what extent do you feel a part of, or close to, this church and its members?" A 6-point graphic rating scale was provided.

2. Acquaintanceship range questionnaire. The Ss were given a list of some members of their church, and were asked to indicate if they (*a*) recognized the name, (*b*) knew the person personally, and (*c*) addressed the person by his first name. Each list contained the names of lay officers in the major positions of the church along with names of randomly selected church members. For the small church, 12 officers and 17 other members (5% of the membership) were listed. For the large church, 16 officers and 80 other members (also 5% of the membership) were named. The same functions were served by the officers listed for the two churches; the large church had more specialized offices in several cases.

3. Participation questionnaire. The questionnaire contained a list of all the activities which had occurred in each church during the 15-wk. period preceding the testing sessions. The *S*s checked the activities they had attended; indicated the number of hours spent in each; and noted whether they had held a major responsibility (leader category), a minor responsibility (worker), or merely attended (member). (See Barker, 1968, Barker & Gump, 1964, and Wicker, 1969b, for a fuller description of the responsibility categories.)

Role-Playing Subjects and Procedure

In an attempt to learn if there is a generally held view that church size is related to new members' behaviors and experiences in church activities, the present questionnaires were also administered to college students who were asked to role play being a new member in either a large or a small church. Two forms of the questionnaire, one for the large church and one for the small church, were prepared for the role-playing *S*s. Each form was completed by 21 undergraduate students at the University of Illinois.

Information about the church and the average new member was provided in the instructions. To illustrate, here is the description of the small church:

> Imagine that you are married and have two children. About 2 yr. ago, you moved to an industrial suburb of a metropolitan area of approximately one million people. About a year ago, you and your spouse joined a nearby Methodist church having around 300 members. The church pastor is well liked by the church members, and he is known as being a good speaker. He is in his 60s, and has been assigned to the church for 2 yr. The present church building was constructed several years ago.

Both forms contained identical questions on assimilation taken verbatim from the new-member questionnaire. Questions on acquaintanceship and participation differed on the two forms, as they did on the original questionnaires. To obtain information on acquaintanceship range, *S*s were told to suppose they had been given a list of 12 (small church form) or 16 (large church form) church officers, and then to estimate the number of these officers whose names they would recognize, whom they would know personally, and whom they would address by the first name. The same questions were asked about members, with either 17 (small church form) or 80 (large church form) being the number of members specified. Data on participation in church activities were obtained by asking *S*s how many of the 42 (small church form) or 104 (large church form) activities occurring in the church during a 3½-mo. period they would have attended, the number of hours they would have spent in the activities during the period, and the number of activities in which they would have had a responsible role.

Analysis of Questionnaire Data

A preliminary correlational analysis on data from the church members indicated that all of the present assimilation measures were positively correlated, with the highest coefficients in the .80s. It was therefore decided to treat the various measures as items in an overall test of assimilation of new members by calculating composite scores on which group comparison would be made. Scores on each of the 12 assimilation measures listed in Table 1 were converted into standard scores. Composite assimilation scores were obtained by summing across the 12 standard scores for each *S*.

Two *t* tests and a 2 × 2 analysis of variance were employed to compare groups on assimilation. One *t* test compared the new members of the large church and the small church, while the other compared *S*s in the large church and the small church role-playing conditions. The analysis of variance tested the main effects of type of *S* (church member or role player) and church size (small or large) and the interaction of the two variables. Composite assimilation scores were calculated separately for each analysis using standard scores based on data from all *S*s involved in the comparison. For example, in the comparison of members of the small and large churches, the standard scores making up the assimilation composite were computed using the means and standard deviation of the distributions for members of both churches. Thus, composite scores indicated degree of assimilation of each *S* relative to all church members, and not just relative to members of his own church. For the analysis of variance, the distributions used for calculating scores were from data for all *S*s (both church members and role-playing *S*s).

Since complete data were required to sum across the measures, it was necessary to drop from the analysis several *S*s who had not completely filled out the assimilation questions. Of the church members, 5 from the small church and 4 from the large church were eliminated, leaving *n*s of 21 and 36, respectively. Of the role-playing *S*s, one small church *S* and one large church *S* were dropped, leaving *n*s of 20 for each condition. Examination of the data suggested that elimination of these *S*s had a negligible effect on group means. It was also necessary to omit the measure of feeling welcome, since a number of church members who had not attended the activities did not indicate how welcome they felt to attend.

Data from Church Records

Frequency of members' attendance at Sunday worship service was determined from church records; in each church a sign-up sheet is passed along the pews during each service. Also available from church records were the amounts of money pledged and contributed to the church by *S* and his spouse (a combined figure for both) during the fiscal year in which the study was conducted.

Comparison of New Members and Established Members

Data from the earlier study of members of the two churches (Wicker, 1969b) were employed in the analysis of the effects of church size, recency of membership, and their interaction on support for church activities. In an attempt to make the samples of the two studies more comparable, the following steps were taken: (*a*) new members in the present study who did not have a minor child were eliminated, since all *S*s in the prior study had at least one such child, and (*b*) *S*s in the earlier study who had been members for 30 mo. or less when they provided data were assigned to the new member category, leaving in the established group only persons who had been members longer than 30 mo. The *n*s and mean length of membership for the resulting four groups were as follows: small church, new members, $n = 30$, $\overline{X} = 1.3$ yr.; large church, new members, $n = 37$, $\overline{X} = 1.4$ yr.; small church, established members, $n = 18$, $\overline{X} = 7.4$ yr.; large church, established members, $n = 33$, $\overline{X} = 11.8$ yr.

New and established members were compared on their participation in church activities, Sunday worship service attendance, pledges, and contributions. A limitation of the comparisons is that the measures on most of the new members (all except those from the earlier study) were obtained 1 yr. after the same measures on the established members. The single exception is data on members' pledges, which are for the same period. Both churches are relatively stable, however, and identical periods of time were used for the 2 yr.

TABLE 1

MEAN SCORES ON ASSIMILATION MEASURES FOR CHURCH MEMBERS AND ROLE-PLAYING *S*s

Measure	Church members		Role-playing *S*s	
	Small church	Large church	Small church	Large church
No. of invitations to join or attend three activities	5.7	5.5	8.3	8.2
No. of offers to take *S* to three activities	3.0	1.4	5.8	7.2
Frequency of attendance at three activities	4.7	1.6	5.8	8.1
No. of times *S* has asked others to attend three activities	2.8	.8	4.0	4.7
Felt obligation to participate in three activities (larger no. indicates greater felt obligation)	6.6	5.4	6.1	7.5
Felt obligation to support four church activities ("yes" answers to four questions)	2.7	2.0	2.7	2.2
Sense of belonging to the church (1 = not at all a part, 6 = very much a part)	4.2	3.6	4.5	4.2
No. of church officers known personally or addressed by first name	5.1	1.6	6.2	5.6
No. of church members known personally or addressed by first name	1.0	2.6	8.7	19.5
No. of activities attended	3.6	2.8	14.6	18.4
No. of hours spent in activities	16.5	14.1	29.2	24.5
No. of activities in which *S* was a worker or leader	2.0	.8	7.7	6.7
Composite assimilation score based on the 12 measures: *t* tests	3.0[a]	−1.7[a]	−.9[b]	.9[b]
Composite assimilation score based on the 12 measures: *F* test	−1.0[c]	−5.1[c]	3.9[c]	6.2[c]

Note.—For church members in small church, $n = 21$; in large church, $n = 36$. Both groups of role-playing *S*s had *n*s of 20.

[a] *t* for difference between means for members of small and large churches $= 2.14$, $df = 55$, $p < .05$ (two-tailed).

[b] *t* for difference between means for role playing *S*s in small and large church conditions $= .70$, $df = 38$, *ns*.

[c] *F*s for analysis of variance: church size, $F < 1$, $df = 1, 93$, *ns*; type of *S*, $F = 28.47$, $df = 1, 93$, $p < .01$; interaction, $F = 4.51$, $df = 1, 93$, $p < .05$.

RESULTS

Group means for the 12 assimilation measures and for the composite scores are shown in Table 1. As expected, new members from the small church were significantly higher on the assimilation composite than new members from the large church ($t = 2.14$, $df = 55$, $p < .05$). Moreover, there was no difference between the mean assimilation composite scores for the large and small church role-playing conditions ($t = .70$, $df = 38$, *ns*), suggesting that the present results for church members are not due to any widely held view of the effects of organization size on members' experiences.

All four groups were compared on the assimilation composite using an unweighted means analysis of variance for unequal cell frequencies (Winer, 1962, pp. 222–224). There was no overall effect on church size ($F < 1$), but the type of *S* main effect was highly significant ($F = 28.47$, $df = 1$, 93, $p < .01$), with role-playing *S*s reporting greater assimilation than church members. The interaction of church size and type of *S* was also significant ($F = 4.51$, $df = 1$, 93, $p < .05$). Figure 1 shows that assimilation reported by church members is greater in the small church than in the large church, but for role-playing *S*s, greater assimilation is reported in the large church condition. In summary, the analysis of variance suggests that the role-playing *S*s were not able to duplicate either the overall level of assimilation of new members or the direction of the difference in assimilation of members of the small church and the large church.

Examination of the means for the individual assimilation measures gives essentially the same picture as the overall analyses. Means for members of the small church were higher than corresponding large church means for 11 of 12 measures. The single exception was number of church members known per-

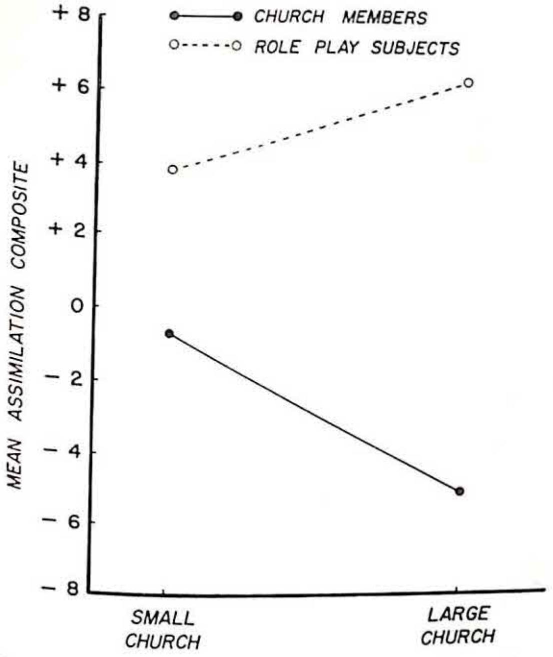


Fig. 1. Mean assimilation scores for church members and role-playing *S*s.

sonally or addressed by the first name, and this result is somewhat ambiguous since members of the small church were given a much smaller list of names (17, compared to 80 for the large church). Role-playing *S*s in both conditions generally had higher scores than the church members, and the small church mean was higher than the large church mean for only 6 of the 12 measures.

Comparative data on support for church activities by new and established members in the large church and the small church are reported in Table 2, along with results of unweighted means analyses of variance. Consistent with earlier research (Wicker, 1969b), church size is significantly related to the range of activities entered ($p < .01$), responsibilities taken ($p < .05$), worship service attendance ($p < .01$), pledges ($p < .01$), and contributions ($p < .06$). For each measure, greater support is shown by members of the small church. The table also shows that new members attend fewer activities ($p < .01$), and a narrower range of activities ($p < .01$), serve in fewer positions of leadership ($p < .05$), and attend worship service less often ($p < .10$) than persons who have been members for more than 30 mo. For none of the measures is the interaction of size of church and length of membership significant, suggesting that the difference between new and established members' support for church activities is the same in the large church and the small church.

Discussion

The present results on new members' assimilation into the two churches supports the hypothesis derived from behavior setting theory that assimilation is facilitated in organizations whose activities are relatively undermanned, that is, small organizations. When compared on composition assimilation measures based on reported participation in church activities, felt obligation to participate, felt closeness to the church, and range of acquaintance with other members and officers, the members of the small church showed greater assimilation than the members of the large church.

The fact that college students role playing members of small and large churches showed no difference in assimilation tends to discount alternative explanations of the obtained differences for church members, such as demand characteristics, questionnaire wording and format, and culture-wide views about the effects of organization size on members' experiences.

The factorial analysis of members' support for church activities generally indicated

TABLE 2

Comparison of New and Established Members in a Large Church and a Small Church

Measure[a]	Small church[b]		Large church	
	New members[c]	Established members	New members	Established members
1. No. of activities attended	5.5	10.2	3.4	7.5
2. No. of kinds of activities attended	4.6	6.4	2.4	4.7
3. No. of activities in which *S* was a worker or leader	2.8	4.6	1.1	2.8
4. Sunday worship service attendance	19.9	26.2	14.1	14.5
5. Pledged contribution ($)	173	258	145	141
6. Actual contribution ($)	126	187	109	109

Note.—In small church, new members, $n = 30$; established members, $n = 18$. For large church, new members, $n = 37$; established members, $n = 33$.

[a] All interactions were *ns*.

[b] Significance levels (two-tailed) for small versus large churches were: $p < .01$ for Measures 2, 4, and 5; $p < .05$ for Measure 3, and $p < .06$ for Measure 6; Measure 1 was *ns*.

[c] Significance levels (two-tailed) for new versus established churches were: $p < .01$ for Measures 1 and 2; $p < .05$ for Measure 3; and $p < .10$ for Measure 4; Measures 5 and 6 were *ns*.

greater support by members of the small church and by persons who have been members longer than 30 mo. The significance of the effects of church size is consistent with earlier research supporting behavior setting theory (Baird, 1969; Barker & Gump, 1964; Wicker, 1968, 1969a, 1969b; Willems, 1967). The relationship of recency of membership and support may be at least partially due to the fact that not all new members remain in the church; some become inactive and drop out. Thus, the established members belong to a more select group which has shown sufficient interest to remain in the church longer than 30 mo. Also, the established members are probably less geographically mobile than new members.

A question can also be raised as to whether new members select churches of the size which makes demands consistent with their willingness to expend effort, since they tend to show the same church size differences in support as do established members (see Table 2). Data collected in the present study but not reported above indicate that new members of large and small churches agree that members of small churches, relative to members of large churches: (*a*) are more friendly, (*b*) spend more time in church activities, (*c*) are more regular in church attendance, (*d*) work harder toward church goals, and (*e*) have a closer relationship with the pastor. They also agree that in large churches, church schools are better and that there is more formality. These data suggest that church members have accurate perceptions regarding correlates of church size (cf. Wicker, 1969b).

In spite of these beliefs, however, new members of both churches rated church size as relatively unimportant in influencing their decisions to affiliate with the present church. Members of the small church rated church size tenth in importance of 13 items; only the music program, availability of parking near the church, and the social standing of the church and its members were rated less important. Members of the large church rated church size twelfth in importance; only availability of parking was judged less important. (The minister was rated by members of both churches as the most important factor, followed, respectively, by location of the church, friendliness of members, worship service, church school, youth program, church facilities, involvement of the church in social issues, and social activities.) Also relevant to the selectivity question is the fact that 42% of the members of the small church and 25% of the members of the large church reported visiting another church before deciding to join their present church. In sum, while beliefs regarding the correlates of church size probably do have some influence on the choice of a church, the immediate situation of undermanning or overmanning is undoubtedly also a critical determinant of levels of support.

Some unanswered questions regarding the influence of church size on members' levels of participation and the assimilation of new members may be clarified by a study in progress on the merger of a church having 400 members and a larger church of 950 members. Data on levels of participation before and after the merger and feelings of being welcome and obligated to participate will permit a test of the effects of church size largely uncontaminated by the selective choice factor.

REFERENCES

BAIRD, L. L. Big school, small school: A critical examination of the hypothesis. *Journal of Educational Psychology*, 1969, **60**, 253–260.

BARKER, R. G. Ecology and motivation. *Nebraska Symposium on Motivation*, 1960, **8**, 1–50.

BARKER, R. G. *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford: Stanford University Press, 1968.

BARKER, R. G., & GUMP, P. V. *Big school, small school: High school size and student behavior*. Stanford: Stanford University Press, 1964.

WICKER, A. W. Undermanning, performances, and students' subjective experiences in behavior settings of large and small high schools. *Journal of Personality and Social Psychology*, 1968, **10**, 255–261.

WICKER, A. W. Cognitive complexity, school size, and participation in school behavior settings: A test of the frequency of interaction hypothesis. *Journal of Educational Psychology*, 1969, **60**, 200–203. (a)

WICKER, A. W. Size of church membership and members' support of church behavior settings. *Journal of Personality and Social Psychology*, 1969, **13**, 278–288. (b)

WILLEMS, E. P. Sense of obligation to high school activities as related to school size and marginality of student. *Child Development*, 1967, **38**, 1247–1260.

WINER, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

(Received January 19, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 157–160

DOUBLE CROSS-VALIDATION OF A WEIGHTED APPLICATION BLANK OVER TIME

DARRELL E. ROACH[1]

Nationwide Insurance, Columbus, Ohio

This research is concerned with a second cross-validation of a weighted application blank for predicting tenure of clerical employees. The sample consisted of employees hired during a later time period than the initial and cross-validation samples. The blank was developed and cross-validated in a previous study. Application blanks for 154 clerical employees hired during late 1953 and 1954 were scored using the weights previously developed. A substantial loss in predictive power was found in the additional cross-validation. Changes in labor market conditions, manpower needs, and specific changes in personnel policies appear to be responsible for the loss in efficiency. The results suggest cross-validation of such blanks with samples of employees hired during a later time period and periodic checks on their continuing validity.

The utility of the weighted application blank for the identification of probable early terminators among job applicants has been demonstrated across a variety of employee populations. Mosel and Wade (1951) used biographical data to reduce turnover among department store clerks and Kriedt and Gadel (1953) used the technique to predict clerical terminators. Dunnette and Maitzold (1955) and Kirchner and Dunnette (1957) reported successful prediction of early termination among seasonal workers in a cannery and female office employees. Fleishman and Berniger (1961) report successful identification of early terminators among clerical employees in a university setting.

The characteristic model for such studies is to obtain application blank data for employees hired 2–3 yr. prior to the research. The sample is then classified into early terminators, long-term employees, and a middle group who terminated employment prior to the study but were active employees for a longer period than the early terminators. Application blank items are then examined to find those which differentiate between the long- and short-term groups. The middle group is usually dropped from the study. Scoring weights are calculated for the discriminating items and these weights and items are cross-validated either by a hold-out sample or by selecting another sample of employees hired during a different time period. Often this is the year following the initial validation sample.

Since the scoring keys for such blanks are generally developed on a sample of employees hired 2–3 yr. prior to the actual use of the instrument in the selection process, the question may be raised as to whether such instruments may be time sensitive and not be as valid for the present applicant population. A study by Buel (1964) in a 3–5-yr. follow-up found that, in spite of a move of the office, a different applicant population, and a tight labor market, 13 of the original 16 items continued to discriminate and the validity was reduced from .49 to .33. Dunnette, Kirchner, Erickson, and Banas (1960) report continued discrimination for the blank developed in an earlier study. However, a further follow-up of that study by Wernimont (1962) found that after 5 yr. the validity had dropped to near zero and only 3 of the original 16 items held up. Thus, it appears that the research on the stability of the predictive power of the weighted application blank is far from conclusive and could be situational.

The purpose of the present study was to conduct a second cross-validation of a weighted application blank using employees hired during a later period than the sample used in the original validation and cross-validation in order to determine the stability of its predictive power prior to incorporating it into the company's selection procedures. In addition, it was desired to identify specific factors which appear to be responsible for any changes in validity by reviewing personnel policy changes, external factors, etc.

METHOD

The development and cross-validation of the weighted application blank used in the present study is reported by Minor (1958). The sample in this study consisted of 440 female clerical employees in the lower salary grades (2–5) hired during 1952–1953. Only salaried employees were included in order to eliminate temporary summer workers who are employed on an hourly basis. One-half of the sample was used for the development of the prediction key and the other half for cross-validation. The samples were selected so that each consisted of a long-tenure group (tenure ranging from 42–54 mo.) and a short-tenure group (terminated prior to 9 mo). The termination group was limited to voluntary terminations thus excluding discharges and resignations due to pregnancy. The composite correlation between the tenure criterion and the weighted blank was .53 for the initial sample and .46 for the cross-validation sample.

---

[1] Requests for reprints should be sent to the author, Research Service, Nationwide Insurance, 246 North High Street, Columbus, Ohio 43216.

The sample in the present study consisted of 199 female clerical employees hired during late 1954 and early 1955. Only full-time salaried employees on the lower salary grades were included. The application blanks of these employees were scored using a slight modification of the weights developed in the study by Minor (1958). These weights were z scores for the item alternatives and the modification consisted of adding the largest negative z to the weights for a given item, dividing by the smallest positive value after the addition of the constant, and rounding to the nearest integer value. This greatly simplified the scoring and since the z scores are relative, the procedure would have a negligible effect on the relationship to the criterion.

During the period covered by the Minor (1958) study, the company was using the Life Office Management Association test (No. 2 Form A) which was published by the firm of the same name for use by its member companies. The test consisted of six subtests involving number checking, following directions, vocabulary, proverbs and sayings, simple arithmetic, and spelling. The scores were the errors in each subtest and the time required to complete the total test. These scores were then combined into a weighted total score. This total weighted score was the only test score utilized in the weighted blank developed in the prior study.

TABLE 1

DISTRIBUTION OF WEIGHTED APPLICATION BLANK SCORES

Score	Long tenure		Short tenure	
	F	%	*F*	%
27–29	1	1.1	—	—
24–26	4	4.4	1	1.6
21–23	15	16.5	2	3.2
18–20	20	22.0	13	20.6
15–17	24	26.4	23	36.5
12–14	20	22.0	15	23.8
9–11	7	7.7	7	11.1
6–8	—	—	1	1.6
3–5	—	—	1	1.6
n	91	100.1	63	100.0
M	17.10		14.81	
σ	3.97		3.54	

Employees hired during the period sampled in the present study were tested with the Wonderlic personnel test. Since the item content of these tests appeared very similar, it was assumed, perhaps naïvely, that the tests would be fairly equivalent and the Wonderlic scores could be used for the purposes of the present study. The weight for the Wonderlic scores was derived from the Life Office Management Association score, assigning comparable percentile ranks.

The present sample was classified into three subsamples. There were 63 employees who terminated prior to 9 mo, 45 who terminated between 10–23 mo., and 91 who were still employed at the time of the study, thus having 23–29 mo. of service. Only the long- and short-tenure groups were used in the analysis, and discharges and resignations due to pregnancy were excluded from the short-tenure group in order that the data would be comparable to the previous study.

The means and standard deviations of the total weighted score and each of the weighted item scores were computed. Point biserial correlations were also calculated for the total weighted score and each weighted item with the criterion.

RESULTS

Table 1 shows the distribution of the weighted application scores for the long- and short-tenure groups. The scores for the long-tenure group range from 9–27 with a mean of 17.10; for the short-tenure group they range from 5–24 with a mean of 14.81. A t test of the difference between the means is 3.75 which is significant at the .001 level. The point biserial correlation between the scores and the criterion is .29. This is substantially below the .46 found in the initial cross-validation.

Examination of the distributions of the scores indicates considerable overlap, with almost complete overlap in the low score range and less overlap in the higher range of scores. A total of 22% of the long-tenure group scored above 20 compared to 4.8% for the short-tenure group. At the other end of the distribution, however, 7.7% of the long-tenure group had scores below 12 compared to 14.3% for the short-tenure group.

Table 2 shows the means for each of the weighted items and the tests for the significance of the difference between means. In addition, a chi-square test was calculated to test for differences in the responses to the item categories since this test would be independent of the weights used. Four of the 11 weighted items show significant discrimination between the long- and short-tenure groups, but two of the items show a reversal of direction. Only two of the items show a significant difference

TABLE 2

ITEM MEANS, DIFFERENCES, t, AND CHI- SQUARE TESTS OF SIGNIFICANCE

Item no.	Item	$\bar{X}$ active[a]	$\bar{X}$ terminated[b]	MA-MT diff.[c]	t	χ^2	df
1	Age at time of application	1.12	.78	.34	1.81	4.71	3
2	Distance of home address from company	2.10	1.79	.31	1.18	4.47	4
3	Length of time married at time of application	1.76	1.46	.30	2.43**	10.61**	3
4	Difference between salary expected & salary received	1.68	1.70	−.02	.19	.81	4
5	Source of reference to company	3.35	2.05	.30	1.41	2.89	8
6	Average tenure on last three jobs	1.18	1.48	−.30	2.72***	7.77	4
7	Reason for leaving last job	1.95	1.51	.44	2.23*	8.28	5
8	Major field in high school	2.80	2.53	.27	1.34	3.33	5
9	Father's occupation (classified by North-Hatt scale)	1.42	1.42	.00	.00	2.59	1
10	Related to or knows company employees	.39	.32	.07	.90	.92	3
11	Wonderlic test score	.58	.81	−.30	2.34*	8.17*	3

[a] n = 91.
[b] n = 63.
[c] $\bar{X}$ active-$\bar{X}$ terminated difference.
* $p = .05$.
** $p = .02$.
*** $p = .01$.

in response categories. Apparently, a weighting procedure which assigns the highest weights to the most discriminating categories achieves significance in some cases where the raw category frequencies do not.

The point biserial correlations of the items with the tenure criterion for the initial validation, and first and second cross-validation samples are shown in Table 3. Almost all of the item-criterion correlations show further shrinkage in the present sample. Items 1, 2, 6, 8, 9, and 10 which held up fairly well on cross-validation showed substantial shrinkage. Only Item 3 appeared to hold up under this second cross-validation. Also, it should be noted that Variables 6, 10, and 11 show a reversal of sign.

DISCUSSION

The weighted application blank developed in the previous study by Minor (1958) discriminates between long- and short-tenure employees in the second cross-validation. The predictive efficiency of the instrument, however, is sharply reduced in the present study. Such shrinkage is to be expected between the validation and cross-validation samples since the initial validation capitalizes on all chance differences. Similar loss in efficiency would not be expected between samples not involved in the determination of the scoring weights.

The overlap of the distributions renders the instrument relatively ineffective for selection purposes. Such instruments are primarily used as rejection devices in that low scoring applicants are rejected since they have a high probability of terminating within a short period after the initial employment. With the overlap at the lower end of the distribution, the probability of early termination of an applicant who scores below 12 is .56 compared to .41 for the total score range of the two samples. At the other end of the distribution, we note that applicants who score above 20 have a probability of .87 of being in the long-tenure group. Thus, it appears that the validity of the instrument in this second cross-validation is primarily associated with this smaller overlap in the high score range. Consequently, it does not appear useful as a rejection device for screening out poor tenure risks.

The examination of the item validities yields some clues as to possible reasons for the loss of predictive efficiency in this second cross-validation. The items concerning age at time

TABLE 3

COMPARISON OF ITEM-CRITERION CORRELATIONS OF VALIDATION AND FIRST CROSS-VALIDATION SAMPLE WITH SECOND CROSS-VALIDATION SAMPLE

Item no.	Item title	Sample		
		1	2	3[a]
1	Age at time of application	.24	.28	.13
2	Distance of home address from company	.21	.17	.09
3	Length of time married at time of application	.31	.21	.19
4	Difference between salary expected and salary received	.18	.07	−.01
5	Source of reference to the company	.34	.15	.13
6	Average tenure on last three jobs	.25	.27	−.16
7	Reason for leaving last job	.25	.10	.16
8	Major field in high school	.28	.20	.11
9	Father's occupation (classified by North-Hatt prestige scale)	.17	.19	.02
10	Related to or knows company employees	.16	.28	−.08
11	Life Office Management Association test score or Wonderlic	.15	.08	−.19

[a] Wonderlic used in Sample 3 on Item 11.

of application, length of time married at time of application, source of reference to company, and reason for leaving last job, held up fairly well in the second cross-validation. The items dealing with distance of home address from Columbus and major field in high school showed some shrinkage but were still in the same direction. The remaining items either failed to discriminate at all or reversed direction.

In the time period between the two studies some personnel policies were changed which could affect the validity of these items. In 1953 a major change was made in the salary schedule in which the beginning salary levels were raised. This could have affected the discriminability of the item dealing with the differential between salary expected and salary received. In 1954 a change was made in the company's relative rule. Prior to that time the company policy prevented the employment of relatives of present employees. This was modified so that female relatives of female employees could be employed. This, of course, has a direct bearing on the item referring to the applicants being related to or knowing a company employee.

The change in the selection test also affected the results. In the previous study it was found that high scorers on the Life Office Management Association test tended to be better tenure risks. With the Wonderlic this tendency was reversed. It was assumed in the present study that these two tests would be highly correlated. Apparently there are differences in the tests or the applicant populations which affect the relationship between test performance and tenure.

Another factor which could be contributing to the loss of predictive efficiency of the instrument is that during the 1952–1953 time period the company was undergoing a decentralization program which required the addition of many clerical employees and it was faced with a tight labor market situation in the area. During the 1954–1955 period this program was completed and the labor market situation was eased somewhat. Thus, the 1952–1953 sample may have been a relatively unselected population compared with the usual employee population.

It appears that the weighted application blank when used to predict probable tenure is sensitive to changes in personnel policies and employment conditions. This suggests that it would be worthwhile to cross-validate such blanks with a sample of employees hired during a different time period than the original validation sample. In addition, the need for periodic monitoring of such instruments to determine loss in predictive efficiency is reaffirmed.

SUMMARY

This study reports a second cross-validation of a weighted application blank using a sample of employees hired during a later time period than the initial and cross-validation samples. There was a large decrement in the predictive efficiency of the instrument between the initial and second cross-validation samples. Examination of specific items which lost predictive power and review of company history suggests that changes in labor market conditions, manpower needs, and personnel policies may be contributing factors to the loss in efficiency. The results suggest the desirability of cross-validating weighted application blanks with a sample of employees hired during a later time period than the initial validation and reaffirms the need for periodic monitoring of such instruments.

REFERENCES

BUEL, W. D. Voluntary female clerical turnover: The concurrent and predictive validity of a weighted application blank. *Journal of Applied Psychology*, 1964, **48**, 180–182.

DUNNETTE, M. D., & MAITZOLD, J. Use of a weighted application in hiring seasonal employees. *Journal of Applied Psychology*, 1955, **39**, 308–310.

DUNNETTE, M. D., KIRCHNER, W. K., ERICKSON, J., & BANAS, P. Predicting turnover among female office workers. *Personnel Administration*, 1960, **23**, 45–50.

FLEISHMAN, E. A., & BERNIGER, J. One way to reduce office turnover. *Personnel*, 1961, **37**, 63–69.

KIRCHNER, W. K., & DUNNETTE, M. D. Applying the weighted application blank technique to a variety of office jobs. *Journal of Applied Psychology*, 1957, **41**, 206–208.

KRIEDT, P. H., & GADEL, M. S. Prediction of turnover among clerical workers. *Journal of Applied Psychology*, 1953, **37**, 338–340.

MINOR, F. J. The prediction of turnover among clerical employees, *Personnel Psychology*, 1958, **11**, 393–409.

MOSEL, J. N., & WADE, R. R. A weighted application blank for reduction of turnover in department store sales clerks. *Personnel Psychology*, 1951, **4**, 177–184.

WERNIMONT, P. F. Re-evaluation of a weighted application blank for office personnel. *Journal of Applied Psychology*, 1962, **46**, 417–419.

(Received February 2, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 161–164

SCREENING EMPLOYMENT APPLICANTS FOR ATTITUDES TOWARD THEFT[1]

PHILIP ASH[2]

University of Illinois at Chicago

The Reid Report is a three-part instrument (scoreable attitudes-toward-theft inventory, biographical data blank, admissions list of past delinquencies) designed to predict the likelihood that an employment applicant will steal from his employer. For a sample of 254 employment applicants, Reid Report evaluations were correlated with evaluations based on polygraph (lie detector) interviews. For each device the evaluation alternatives were: recommended for hire, qualified recommendation, not recommended. Chi-square for the 3×3 table was 45.15 ($p < .001$). A Pearson product-moment correlation coefficient between polygraph evaluations and Reid Report evaluations was .43 ($p < .001$).

Employee theft of money and merchandise constitutes a heavy tax on business, industry, and ultimately the consumer. Jaspan (1968) estimated that white-collar workers alone steal more than $5 million a day in cash and merchandise. It is estimated that the loss of stock certificates and securities from the New York Stock Exchange through theft was $12.4 billion in 1968 (Anonymous, 1969). Police-reported losses in 1968 for shoplifting were $504 million, and $381 million for employee theft (Statistical Abstract of the United States, 1969). These are undoubtedly gross underestimates. Martin (1970), Chairman of the Board of Carson, Pirie, & Scott (a Chicago department store chain) reported that a survey of retail department stores in that city showed shrinkage (unexplained disappearance of money and merchandise—primarily a euphemism for theft by employees and shoppers), which ranged from 1.6% to 4.7% of gross sales, an amount typically greater than before-tax profits.

Attempts to control and reduce losses due to theft, particularly employee theft, include three principal techniques: apprehension and prosecution of the thief, institution of security procedures and devices to make theft difficult or its apprehension easier, and screening of employment applicants in an attempt to identify those who might steal if employed.

As seems to be true in the case of most other crimes, apprehension and prosecution has very little deterrent effect. If fact, an offender is apprehended in less than 20% of the crimes that are committed (Faltermayer, 1970).

Security devices also proliferate, but apparently only with minor effect. What the mind of one man can invent, the mind of another can evade.

The screening of employment applicants is the third line of defense. Several screening devices have been utilized.

In high security and high trust positions (e.g., bank tellers and bank guards, armored car money transfer personnel, police departments) the polygraph (or lie detector) is frequently used for screening applicants. The device has been demonstrated to possess a high degree of reliability (Bersh, 1969; Horvath & Reid, 1971; Kubis, 1962; Reid, 1971) in establishing an applicant's innocence or guilt with respect to specific criminal activities. The procedure is relatively expensive, however, and unions have strongly resisted its routine use in employment selection. Eleven states now have antipolygraph laws: Alaska, California, Connecticut, Delaware, Hawaii, Massachusetts, New Jersey, Oregon, Pennsylvania, Rhode Island, and Washington (Reid & Hunter, 1970).

A second source of information is the reference, credit, and police check. At best, however, this check is incomplete: it never detects the previously unapprehended. At worst, it is deceptive: ex-employers are too frequently unwilling to commit themselves in a written reference to an allegation that a previous employee stole from them. The possible costs in libel suits far outweigh the limited return earned by candor.

---

[1] Grateful acknowledgment is made to John E. Reid and Associates, and particularly to John Reid, President and to J. Kirk Barefoot, formerly Director of Security, McKesson and Robbins, for making the Reid Report and polygraph data available for analysis.

[2] Requests for reprints should be sent to the author, Department of Psychology, University of Illinois, Box 4348, Chicago, Illinois 60680.

A third source of information is the test or inventory that might predict delinquency. Significant predictions have been claimed for a wide variety of instruments. Among them are Activity Vector Analysis (Clark & Hasler, 1967), the Minnesota Multiphasic Personality Inventory (Hathaway & Monachesi, 1953), the Porteus Mazes (Porteus, 1954), the Id-Ego-Superego Test (Mangold, 1966), the Rorschach (Ichimura, 1966), knowledge of criminal argot or vocabulary (Gibson, 1966), the Kvaraceus Delinquency Scale and Checklist (Bechtold, 1964; Kvaraceus, 1954), the California Personality Inventory (Gough & Peterson, 1952), the Reid Report (Ash, 1970; Reid, 1967), and a variety of other devices too numerous to list.

The present study is concerned with an application of the Reid Report to the identification of potential thiefs. The Report differs from most previous devices in two respects: first, it has been designed to predict only one kind of delinquent behavior, theft; and second, its purpose should be transparent to the examinee.

METHOD

The Reid Report consists of three main sections. The first section comprises a yes-no questionnaire including two sets of items. The first set yields a measure of punitiveness, as reflected in attitudes toward punishment for crimes of theft (e.g., "Do you believe there are some special cases where a person has a right to steal from an employer?"; "An employer discovers that a long-service, trusted employee has been taking a few dollars out of the cash register every week. Should the employer have him arrested?"). The second set of questions is intended to measure the individual's own attitudes and behaviors relating to theft (e.g., "Did you ever think about committing a burglary?"; "Are you too honest to steal?").

A previous study (Ash, 1970) showed that scores on the inventory had satisfactory validity when correlated with admissions of previous significant thefts elicited during a polygraph examination. In a sample of 1,519 males—both black and white—a validity of .43 or higher was obtained for each of four age groups (under 20, 21–25, 26–35, 36 and over) for both black and white applicants for positions of trust, frequently involving the unsupervised handling of money (e.g., bank tellers and messengers, police applicants, security guards, dairy route salesmen). In an unpublished study by the author, a cross-validation based upon items selected in the earlier study yielded a validity of .58 with the results of polygraph-determined theft history.

The second section of the Report consists of a detailed biographical data blank covering previous employment; education; personal history (e.g., marital status, addresses over past 10 yr.); financial history and indebtedness; medical and social history (use of alcohol and drugs, psychotherapy, experience with police including questioning about arrest for and conviction of a variety of theft-related crimes, and gambling practices).

The third section includes a list of questions to which a "yes" response constitutes an admission of a committed theft or other defalcation (e.g., "Did you write a check knowing there was not enough money in the bank more than three times?"; "Did you make a false insurance claim for personal gain?").

In the typical employment use of the Report, an applicant completes the form in the employer's personnel office. Part 1 is scored and all three sections are analyzed by a staff member of J. E. Reid and Associates. On the basis of the item score, plus evaluation of responses to the bio-data and the admissions questions, an evaluation of "recommended for employment" or "not recommended" is made. Norms exist for the score itself, but the bio-data and the admissions questions are also taken into account in the final recommendation. Thus, for example, a middling to high score, which might by itself lead to a "recommended" evaluation, may be reversed by a bad debt history, work instability, or numerous admissions. Incredible as it may seem, applicants in significant numbers do admit to practically every crime on the books. Furthermore, on the first part, on both the punitiveness and self-reflective items, "faking good" seems to be rare. In the absence of data, it may be speculated that response to the questionnaire is strongly determined by the individual's own practices: someone who steals approves of punishment only for persons who steal more than he does (punitive items), and sees as admissible more "thoughts about stealing" than more rigidly honest respondents.

These speculations about why people admit to theft are well supported by interview experiences with employment applicants in which their implicit norms concerning other persons' theft patterns and their projections about the interviewer's reaction to a denial of any theft experience are vividly illustrated. For example, on the filter question, "As far as your own honesty is concerned do you think you are [a 7-point scale ranging from far above average to far below average]" many applicants rate themselves average or below, and those who do choose these rating alternatives tend to have lower scores on the inventory itself. On the question "What percentage of employees do you think steal something from their company?" the mean estimate is 10%, but low scorers typically estimate 35–50%. The question, "The total amount of money (that did not belong to me) that I have taken from jobs would be about [can't remember, none, $1 to $5,000 in eight increments]" has elicited, in justification for the response given, the most interesting insights. In the case of one applicant, for example, an initial admission of $600 was erased and changed to $1,000. During the interview, the applicant was asked why he made the change, and whether he still had doubts about the right answer. He responded that (*a*) he knew everyone took some money and the examiner would know he was lying if he claimed never to have taken any himself; and (*b*) upon reflection, figuring a 5-yr. period, $1000 was only $4 per week, and "Doc, isn't that so? Isn't that par for the course? I want to level with you, and you would think I was some kind of a freak if I said I never took anything!" (This was an applicant for a job as bank teller, who was then employed in another bank and looking for a change of scenery.) In short, the applicant who has committed thefts from his employers generally seems to believe that "everyone does it" and that denial of any thefts would be viewed by the examiner as transparent lying. His responses tend to reflect what he considers to be the norms for theft behaviors.

The foregoing attitudinal sets probably militate against "faking good," even though the purpose of the inventory should be obvious to most examinees.

Procedures and sample. McKesson and Robbins, a since-reorganized drug manufacturing company, used both the polygraph and the Reid Report to screen employment applicants. Applicants were interviewed while polygraph tracings were made, at plant locations in 34 cities distributed across the United States. On the basis of the polygraph examination, a security department recommendation—"recommended for hire," "qualified recommendation," or "not recommended"—was made. In addition, all applicants completed a Reid Report, which was sent to the office of J. E. Reid and Associates in Chicago for scoring and evaluation. This evaluation was made completely independently of the polygraph determination, which was, at the time of evaluation, not communicated to Chicago.

In all, Reid Reports were evaluated for 1,022 employment applicants. Polygraph recommendations, however, were submitted for 254 of them.

RESULTS

A comparison of the distribution of Reid Report recommendations for cases for which polygraph examinations were available with all 1,022 Reid Report evaluations failed to reveal any significant differences.

TABLE 1

REID REPORT EVALUATIONS FOR ALL CASES EVALUATED AND FOR CASES FOR WHICH POLYGRAPH EVALUATIONS WERE AVAILABLE

Cases	Evaluation			
	Recommended	Qualified recommendation	Not recommended	Total
Reid Report + polygraph findings				
N	156	18	80	254
%	61.4	7.1	31.5	100
All Reid Reports				
N	616	92	314	1022
%	60.3	9.0	30.7	100

TABLE 2

ASSOCIATION BETWEEN REID REPORT EVALUATIONS AND POLYGRAPH EVALUATIONS FOR 254 EMPLOYMENT APPLICANTS

Reid Report evaluation	Polygraph evaluation							
	Recommended		Qualified		Not recommended		Total	
	N	%	*N*	%	*N*	%	*N*	%
Recommended	129	50.8	12	4.7	15	5.9	156	61.4
Qualified	13	5.1	4	1.6	1	.4	18	7.1
Not recommended	35	13.8	13	5.1	32	12.6	80	31.5
Total	177	69.7	29	11.4	48	18.9	254	100

Note.—$\chi^2 = 45.15$, $df = 4$, $p < .001$; $r = .43$, $p < .001$.

For the sample of 254 cases for which both polygraph-based recommendations and Reid Report evaluations were available, a high degree of relationship was found between them (see Table 2).

Chi-square was 45.15 ($df = 4$), significant beyond the .001 level of confidence (Peters & Van Voorhis, 1940, p. 498). Since the recommended–not recommended judgment may be considered a continuous variable, a Pearson product-moment correlation, corrected for broad categories (Peters & Van Voorhis, 1940, p. 395–399), was computed at $r = .43$.

The hit ratio (same recommendation for both polygraph and Reid Report was 65%; the miss ratio (recommended vs. not recommended) was 19.7%. Qualified recommendations, either from the polygraph data or from the Reid Report evaluations, accounted for the balance of the cases, 15.3%.

Furthermore, analysis of the misses indicates that de facto agreement was greater than the statistical tally revealed.

For example, among the 15 misses not recommended (polygraph), recommended (Reid Report), six applicants earned passing scores on the Reid Report and no theft evaluations on the polygraph, but the polygraph examiner rejected them for other reasons, such as medical, heart condition, heavy drinker, wanted sales job—none open, chronic absenteeism.

On the other hand, among those applicants who passed the polygraph examination but were not recommended on the basis of their Reid Report responses, the Reid Report almost always elicited both (*a*) a low score on the first part, and (*b*) admission on the last part of the report of past thefts of various sorts, or (*c*) poor histories including many changes of jobs and home addresses, and credit indebtedness which the applicant could probably never liquidate.

In summary, the Reid Report gives promise of being a device which can identify, at a reasonable level of confidence, applicants who probably have a previous history of theft behavior, and who might be presumed to have attitudes conducive to stealing from subsequent employers, given the opportunity.

REFERENCES

ANONYMOUS. Thievery at a crisis level. *Burroughs Clearing House*, 1969, **53**, 16.

ASH, P. The validation of an instrument to predict the likelihood of employee theft. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 1970, **5**, 579–580. (Summary)

BECHTOLD, M. L. Validation of the K. D. scale and checklist as predictors of delinquent proneness. *Journal of Experimental Education*, 1964, **32**, 413–416.

BERSH, P. J. A validation study of polygraph examiner judgments. *Journal of Applied Psychology*, 1969, **53**, 399–403.

CLARKE, W. V., & HASLER, K. R. Differentiation of criminals and non-criminals with a self-concept measure. *Psychological Reports*, 1967, **20**, 623–632.

FALTERMAYER, E. K. Some here-and-now steps to cut crime. *Fortune*, 1970, **83**(1), 94–97.

GIBSON, H. B. The validation of a technique for measuring delinquent association by means of vocabulary. *British Journal of Social and Clinical Psychology*, 1966, **5**, 190–195.

GOUGH, H. G., & PETERSON, D. R. The identification and measurement of predispositional factors in crime and delinquency. *Journal of Consulting Psychology*, 1952, **16**, 207–212.

HATHAWAY, S. R., & MONACHESI, E. D. (Eds.) *Analyzing and predicting delinquency with the MMPI*. Minneapolis, Minn.: University of Minnesota Press, 1953.

HORVATH, F. S., & REID, J. E. The reliability of polygraph examiner diagnosis of truth and deception. *Journal of Criminal Law, Criminology, and Police Science*, 1971, in press.

ICHAMURA, J. Ten year follow-up study on the early prediction of juvenile delinquency by means of the Rorschach test. *Japanese Psychological Research*, 1966, **8**(4), 151–160.

JASPAN, N. White collar thefts at $5 million a day. *The National Underwriter*, 1968, **72**, 16.

KUBIS, J. F. *Studies in lie detection, computer feasibility consideration*. Griffen Air Force Base, New York: Rome Air Development Center, Air Force Systems Command, United States Air Force, 1962.

KVARACEUS, W. C. *The community and the delinquent*. Yonkers-on-Hudson, New York: World, 1954.

MANGOLD, K. M. Comparison of delinquents and nondelinquents on the IES test. *Perceptual and Motor Skills*, 1966, **22**, 817–818.

MARTIN, V. *A golden past—The challenge ahead*. Speech before the Controllers Congress Fiftieth Anniversary Program, Chicago, May 1970.

PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.

PORTEUS, S. D. *Q*-scores, temperament, and delinquency. *Journal of Social Psychology*, 1945, **21**, 81–103.

REID, J. E. *The Reid Report*. Chicago: Reid, 1967.

REID, J. E. Judicial recognition of the polygraph ("lie detector") technique. *Trial Judges Journal*, 1971, in press.

REID, J. E., & HUNTER, F. L. Eliminating your silent partners. *American Automatic Merchandiser*, 1970, (April), 25–28.

Statistical Abstract of the United States. Washington, D.C.: Government Printing Office, 1969. (Table 214, p. 144).

(Early publication received July 29, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 165–169

SELECTION AND EVALUATION PROCEDURE FOR OPERATING ROOM PERSONNEL[1]

URY GLUSKINOS[2] AND THOMAS F. BRENNAN

Temple University

Sixteen surgeons and supervisors of operating room (OR) personnel rated 100 behavioral statements pertaining to OR performance on a 1–7 importance scale. Forty-two statements were retained and incorporated into a final criterion list on the basis of mean rated importance and agreement among raters as measured by their standard deviations. In the second phase of the study, 163 OR personnel were rated by their supervisors on the 42-item criterion list and of these, 40 were administered a series of tests consisting of the Wonderlic, O'Connor Finger Dexterity Test, Press Test, and three scales of the California Psychological Inventory. Factor analysis of the criterion variables revealed that 54% of the variance could be accounted for by four factors: general technical knowledge, attention to detail, patient awareness, and social interaction. The first three factors were significantly related to one or more of the predictor variables.

In recent years attention has been focused on the inadequacy of the criterion in most prediction studies. Nevertheless, even today, most prediction studies approach the criterion problem in a most superficial way, leading to criterion contamination and limiting its reliability. Exceptions are such studies by Hemphill (1963) on engineers and Smith and Kendall's (1963) study on nurses. The study described here outlines a systematic approach to the criterion problem in a little explored vocational specialty, that of the operating room (OR) personnel.

For many years the demands of hospitals for qualified OR personnel has greatly exceeded the supply of such personnel available. Until approximately 10 yr. ago, ORs have been staffed primarily by graduate and registered nurses who had undergone intensive training in OR procedures. Since nursing schools screen applicants prior to admission and subject their nursing students to rigorous examinations in theoretical knowledge and clinical performance prior to graduation, no great need has existed for a specific OR selection procedure.

However, in recent years, hospital nursing schools have reduced or omitted OR training from their curriculum and many hospitals have begun to initiate training programs for nonprofessional OR personnel (i.e., the surgical technicians). It might be noted that the armed forces have employed corpsmen in their ORs to function as scrub and circulating surgical technicians. Upon discharge from the service, these men and women have demonstrated to hospitals the value of the nonprofessional, functioning as members of the surgical team (Peers et al., 1966). Hospitals are now recognizing that the need to staff their ORs can now be partially satisfied by employing technicians to assist the registered nurse (RN) in the OR.

Individual hospitals construct their own training programs which vary in length from 6 wk. to 18 mo. and selection of applicants is unique for each hospital. Applicants to these programs are primarily premedical students, licensed practical nurses (LPN), women who have been unable to complete their nursing training, auxiliary hospital personnel (orderlies, dietary personnel, etc.) who have expressed an interest in such a program, and in some cases, individuals with no prior medical experience, who have a sincere interest in medicine and were attracted to the program (Boyer, 1966). Selection procedures have not seemed effective according to the high attrition rate as reported by some hospitals (Turkenkopf, 1967). It was with this thought that this study was undertaken.

The objectives of this study were to develop a testing procedure for the selection of OR personnel and to construct an adequate standardized criterion for the evaluation of OR personnel, the necessity of which is recognized in the field of hospital management (Metzger, 1967).

---

[1] This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Ury Gluskinos, who is now at the Israel Institute for Productivity, P.O. Box 33010, Tel Aviv, Israel.

The construction of a standardized evaluation form could be undertaken with the knowledge that OR tasks are essentially the same among the different types of hospitals in the United States, much the same as surgical procedures are performed in the same manner, regardless of the location or type of hospital.

METHOD

Criterion Development

A list of 100 descriptive behavioral statements applicable specifically to performance in the OR was collected. The statements were submitted by OR nurses, OR supervisors, and surgeons and based on actual observation of OR personnel performing these tasks during surgical procedures. These 100 statements were then submitted to 16 surgeons and OR supervisors from various hospitals, and were rated by them on a scale of 1–7 (least important–most important, respectively). Mean importance ratings and standard deviations were computed for each of the statements. Only those statements with means above 3.80 and standard deviations below 1.75 were used as criterion variables. The final evaluation form consisted of 42 statements with a rating scale of the form:

	1	2	3	4	5
Statement	Rarely	Seldom	Frequently	Often	Always

The order of the rating scale was randomly alternated to reduce response sets.

Predictor variables were chosen on the basis of a limited testing time availability and their appropriateness to the types of statements contained in the final evaluation form. The tests selected for use in this study were the: (*a*) O'Connor Finger Dexterity Test; (*b*) Wonderlic Personnel Test; (*c*) three scales from the California Psychological Inventory (CPI; Responsibility, Self-Control, and Socialization); and (*d*) Press Test which yields five scores. The Press Test is a timed paper and pencil test divided into three parts. In the first part, the testee is to write the first letter of a series of words (red, blue, green, yellow; randomly alternated) in the corresponding circle to the right of the words, the words being printed in black ink. The score is the number of circles completed in 90 sec. In Part 2, *S*s must place the first letter of the color of a dot (red, green, blue, yellow) in the appropriate circle. In Part 3, the word of the color is printed in a different colored ink, for example, red printed in green ink, blue printed in yellow ink etc., and again the order is randomly alternated. The *S*s are required to place the first letter of the color of the ink in the appropriate circle to the right of the series of colored words. The stress measure is the score of Part 3, that is, *S*s' ability to discriminate effectively in the presence of distracting stimuli. A total of 10 test scores was obtained.

Procedure. The evaluation form was mailed to 100 hospital OR supervisors in seven states and the District of Columbia. Supervisors from 17 hospitals responded by completing evaluations on 163 OR personnel. This sample consisted of 76 RNs and 87 OR technicians (ORT). Of the 17 hospitals responding, the seven hospitals from the Philadelphia area agreed to have their personnel tested with the predictors chosen for this study. The four tests were administered by the second author to 40 OR nurses and ORTs drawn from these seven hospitals. The sample consisted of 38 females and 2 males whose ages ranged from 19 to 54 yr. ($\bar{X}$ age = 28.4 yr.) with an average of 4.2 yr. experience in the OR.

RESULTS

Analysis of the Criterion Variables

A principal component factor analysis was performed on the evaluation scores of the 163 RNs and ORTs and the resulting factors were rotated to simple structure. Four factors which accounted for 54% of the total variance were retained. These factors lend themselves to fairly unambiguous interpretation; they defined specific areas of OR task performance. Below follows a description of these factors and the major defining statements, each statement being followed by its factor loading and mean importance rating. For each factor, internal consistency reliability coefficients were also computed.

Factor 1—General Technical Knowledge

This factor contains statements which measure an individual's performance of routine OR procedures, such as, application of principles of sterile technique, preparation for surgical procedures as to type of instruments and equipment required, and knowledge of the sequence of events for a particular operating procedure, etc.

Sees that suction machine is functioning properly (.74, 6.06)
Anticipates the needs of the surgeon (.73, 6.37)
Is able to thread needles quickly (.72, 4.25)

The Spearman-Brown corrected reliability coefficient for this factor was computed to be $r = .96$.

Factor 2—Attention to Detail

The statements with significant factor loadings on this factor measure the individual's performance of specific tasks which are essential to effective performance in the OR. The mean importance ratings of all of the statements are high and have low standard deviations, from which we can infer that surgeons and OR supervisors consider these statements to be of extreme importance in the OR. It is also interesting to note that the mean score of the sample ($N = 163$) on this factor was 28.01 with a standard deviation of 2.89 and a range of 18–30, 30 being the highest score attainable on this factor.

Maintains aseptic technique throughout the procedure (.78, 6.93)
Keeps track of needles on the surgical field (.76, 6.62)
Follows proper surgical scrub technique (.72, 6.56)

The Spearman-Brown corrected reliability coefficient for this factor was computed to be $r = .89$.

Factor 3—Patient Awareness

The statements which make up Factor 3 are primarily patient oriented with an underlying safety factor involved in most of the statements. For example, the first statement listed below for this factor measures the frequency with which a nurse checks her shoes for conductivity. With the use of explosive anesthetic gases in the OR there is a hazard which arises from the production of static electricity on the clothes of OR personnel. OR personnel thus wear shoes which will conduct this static electricity from their bodies through the shoes to a special flooring which is grounded and thereby eliminate the hazard of possible arcing of static which can set off the explosive gases. It is essential that these shoes be checked and cleaned frequently to insure the efficient conduction of this static electricity. This particular statement is particularly patient oriented since the greatest concentration of these explosive gases is centered in and around the patient.

Checks her shoes each day for conductivity (.67, 5.18)
Shows sympathy to her patient (.58, 4.87)
Does not move patient after surgery without the consent of the anesthesiologist (.54, 6.06)

The Spearman-Brown corrected reliability coefficient for this factor was computed to be $r = .82$.

Factor 4—Social Interaction

Factor 4 measures an individual's social behavior in relation to superiors, co-workers, patients, and self. The number of social behavior statements collected and observation of performance demonstrated the close personal contact which exists in ORs. The importance of accounting for a social behavior factor in OR performance is also reflected by the relatively high mean-importance ratings given to these statements by surgeons and OR supervisors.

Shows courtesy to her co-workers (.77, 5.30)
Accepts constructive criticism (.77, 6.18)
Will readily accept new members to the staff (.62, 5.24)

The Spearman-Brown corrected reliability coefficient for this factor was computed to be $r = .81$.

Prediction of Performance

Scores for the four factors plus a total criterion score were computed for each *S* and these scores were then correlated with the 10 predictor scores (see Table 1). The greatest number of significant relationships was found for Factor 2, attention to detail. The highest correlation on this factor was obtained with the Wonderlic ($r = .55$, $p < .01$). It is somewhat surprising that such a significant relationship was found for this factor when one considers the tight distribution of scores as mentioned above. Similar correlations were also found with the Wonderlic and the total criterion score and scores on Factor 1, general technical knowledge ($r = .53$ and $.56$, respectively). The Socialization scale of the CPI

TABLE 1

CORRELATION MATRIX OF TEN TEST SCORES WITH FIVE EVALUATION SCORES

Test	Factor				
	Total	1	2	3	4
O'Connor Finger Dexterity Test	−.23	−.19	−.32*	−.31	−.09
Wonderlic Personnel Test	.53**	.56**	.55**	.19	.29
Press Test					
Part 1	.38*	.35*	.43**	.18	.28
Part 2	.45**	.38*	.51**	.42**	.28
Part 3	.37*	.32*	.46**	.25	.24
Part 1–2	.10	.17	.12	−.21	.09
Part 2–3	.26	.19	.21	.41**	.16
California Psychological Inventory					
Responsibility	.23	.27	.13	−.14	.26
Self-Control	−.07	−.01	−.10	−.34	.04
Socialization	.33*	.37*	.33*	−.03	.24

Note.—$N = 40$.
* $p < .05$.
** $p < .01$.

predicted Factors 1 and 2 but failed to predict the social interaction factor (Factor 4) with significance. Factor 1, general technical knowledge, and Factor 2, attention to detail, were predicted by the Press Test which is a measure of an individual's ability to function effectively under stress conditions. Since effective OR performance depends upon the person reacting to events occurring in rapid succession which often do not follow in the usual or expected sequence, Factor 1 and Factor 2 are validations of the Press Test (see Table 1). The Press Test also predicted Factor 3, patient awareness.

Factor 4, social interaction, could not be predicted from any of the test variables used in this study.

Criterion Analysis

An analysis of variance between RNs' and ORTs' five criterion scores was performed to detect any differences in the rated performance of these two groups.

The results of this analysis indicate that the RNs were rated by their supervisors significantly higher ($p < .01$) than the ORTs on their total score as well as their scores for general technical knowledge, attention to detail, and patient awareness (see Table 2). Since the RNs undergo longer and more intensified training in OR technique as well as other areas of patient care than do the ORTs, one would expect that they would perform at a higher level technically in the OR situation, but not necessarily higher socially. This point is verified by the statistically nonsignificant difference between the two groups on the social interaction factor.

TABLE 2

MEANS, STANDARD DEVIATIONS, AND ANALYSIS OF VARIANCE RATIOS FOR CRITERION FACTORS OF RNS AND ORTS

Factor	Group	$\bar{X}$	*SD*	*N*	*F* ratio
Total score	RN	182.49	15.36	76	20.73*
	ORT	168.58	22.43	87	
	Total	175.06	20.61	163	
General technical knowledge	RN	90.20	9.78	76	22.55*
	ORT	80.82	14.60	87	
	Total	85.19	13.39	163	
Attention to detail	RN	28.93	1.78	76	15.86*
	ORT	27.21	3.40	87	
	Total	28.01	2.89	163	
Patient awareness	RN	19.82	3.12	76	8.50*
	ORT	18.26	3.61	87	
	Total	18.99	3.47	163	
Social interaction	RN	43.53	5.13	76	2.29**
	ORT	42.29	5.38	87	
	Total	42.87	5.28	163	

* $p < .01$;
** *ns.*

DISCUSSION AND CONCLUSION

In the interpretation of the above analysis of variance, a question could possibly be raised that the OR supervisors (also RNs) may have rated their nurses more favorably than their technicians as the result of a bias or halo effect. Also, the fact that RNs are considered professional whereas the ORTs are considered nonprofessional may have contributed to the differences in the ratings. If this were the case, one would logically expect that the social interaction factor would be particularly prone to the halo effect, but this notion was not supported by the analysis of variance. These results, then, lend support to the criterion dimensions as measures of OR performance. Since the statements comprising the criterion dimensions were considered the most important task performed in the OR by OR nurses, supervisors, and surgeons, and discriminated between the better trained nurses and the less trained technicians, we do have some evidence of the relevance of the criterion.

There are additional justifications for the use of the factors as criterion dimensions instead of the separate statements.

1. A correlational analysis revealed that the statements which correlated significantly with the predictors were for the most part the same as those which were contained in the factors.
2. The rating scale for a single statement had only five possible alternatives, thus limiting the variability.
3. The ratings obtained in this sample for any of the statements were rarely the first or second alternative (this was to be expected since this was a concurrent study), further limiting the variability.
4. Statistically and intuitively each factor tends to measure one specific dimension of OR performance.

5. For any psychological or physical measurement, an increase in the number of items (provided those items are similar in content as indicated by large internal consistency coefficients) will increase the reliability of the measure.

The usefulness of behavioral statements as a starting point for criterion development was shown in this study which attempts to measure and predict OR performance. Factor analysis revealed four basic dimensions accounting for 54% of the variance of OR performance. For three of these dimensions, sizable relationships were obtained from a combination of four predictors. As a concurrent validity study, these predictors probably underestimate the magnitude of the correlations. Since the sample involved in the prediction phase of the study was rather small ($N = 40$), replications should be made to cross-validate the predictors used. Additional tests may be used to predict the social interaction factor which could not be predicted with the current battery of tests.

REFERENCES

BOYER, S. B. Key factors in successful program for training OR technicians. *OR Yearbook from Hospital Topics*, 1966, **17**, 39–41.

HEMPHILL, J. K. *The engineering study*. Princeton, N. J.: Educational Testing Service, 1963.

METZGER, N. C. Scientific approach to job evaluation. *Hospital Topics*, 1967, **45**(10), 46–48.

PEERS, J. G., WILKINS, M. E., KUHNS, H. BURNSIDE, F., JOHNSON, H. W., & PRESTI, J. C. Operating-room technicians: Their role and training. *OR Yearbook from Hospital Topics*, 1966, **17**, 28–38.

SMITH, P. & KENDALL, L. H. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 1963, **47**, 149–155.

TURKENKOPF, E. M. Preparing surgical technicians. *Hospital Topics*, 1967, **45**(10), 101–105.

(Received November 25, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 170–174

COMPARISON OF WORK GROUPS WITH STABLE AND UNSTABLE MEMBERSHIP[1]

B. D. FINE[2]

Institute for Social Research, University of Michigan

Work groups with unstable membership are compared to similar groups with stable membership. Contrary predictions concerning the effects of unstable membership were derived from the Likert and the Bennis conceptions of organization structure. Employees in unstable work groups reported less within-group peer leadership, but more favorable organizational context. Contrary to the prediction derived from Likert's theory, there were no differences on measures of group-referrent intervening variables and end-result variables, suggesting the viability of organic-adaptive structure described by Bennis. Increased lateral linkage among work groups of unstable membership is suggested as the mechanism that appears to compensate for the decreased peer leadership evident in unstable work groups.

When an organization abandons traditional structural forms that are readily depicted by simple organization charts, and moves toward a more dynamic and fluid structure, one result may be that work groups with unstable membership become a permanent feature of the new structure. Current theories of organizations disagree about the effects of such unstable structure on organizational functioning. This article contrasts two such theoretical positions and presents empirical evidence relevant to the issue.

Likert (1961) has summarized a considerable body of evidence supporting the conclusion that the quality of interaction process in face-to-face work groups is an important determinant of organizational performance. One principle derived from Likert's theory of organizations is that the superior should supervise his subordinates as a group. The small peer group is the social unit within which supportive relationships and high performance goals can be promoted most effectively. These face-to-face work groups become the component units of an organization. In an organization lacking stable face-to-face work groups, group-oriented methods of supervision would be difficult if not impossible to apply, and interpersonal communication and influence would consequently decline. Capacities for decision making, performance goal setting, motivation, and group loyalty, would decrease, followed eventually by a drop in the levels of end-result variables such as productivity, earnings, satisfaction, and mental health. Indeed, in a detailed list of organizational variables, Likert (1967) included as a causal variable the "extent to which there is sufficient stability in personnel assignments and investment in organization building to create highly effective cooperative working relationships, even though rotation is used for training purposes [p. 218]." These aspects of Likert's theory led to the hypothesis that stable work groups would be higher than unstable work groups on measures of work group functioning, and consequently also on measures of end-result variables.

Bennis' (1966) description of "organic-adaptive" structure, which he sees evolving in organizations that operate under conditions of rapid technological change and growing environmental turbulence, forecasts temporary groupings as a key feature of future organization structure, where "problems will be solved by groups of relative *strangers* [p. 12]." From Bennis' writings was derived the prediction that organizational behavior would not show general impairment where work groups have unstable membership.

---

[1] This investigation was supported in part by Public Health Service Fellowship No. 1 F01 MH 41,073-01 (BEH) from the National Institute of Mental Health. The author is grateful to David G. Bowers, Rensis Likert, and Jon H. Barrett for comments on an early draft of this article, and to Stanley E. Seashore for reading subsequent drafts and providing many helpful suggestions.

[2] Requests for reprints should be sent to the author, now named Dov Eden, presently at the Graduate School of Business Administration, Tel Aviv University, Tel Aviv, Israel.

TABLE 1

MEAN SCORES FOR PLANNABLES AND ASSIGNEDS ON MEASURES OF FOUR PEER LEADERSHIP FACTORS

Peer leadership factor	Work group structure				t
	Plannable		Assigned		
	$\bar{X}$	SD	$\bar{X}$	SD	
Support					
Work group is friendly	3.86	.96	3.95	1.02	
Work group pays attention to you	3.56	.92	3.75	.96	2.04*
Work group willing to listen to you	3.63	1.01	3.70	.99	
Goal emphasis					
Work group encourages effort	3.17	1.12	3.25	1.05	
Work group has high standards	3.66	.91	3.72	.97	
Work facilitation					
Work group shows you how to improve	3.32	1.07	3.59	1.06	2.65**
Work group helps plan ahead	3.25	1.14	3.28	1.26	
Work group gives you new ideas	3.49	1.03	3.50	1.12	
Interaction facilitation					
Work group encourages teamwork	3.22	1.18	3.46	1.13	2.19*
Work group emphasizes team goal	2.94	1.47	3.19	1.10	2.12*
Work group exchanges opinions	3.19	1.06	3.41	1.08	2.12*

Note.—The N is roughly 140 plannables and 525 assigneds. The N's for particular measures vary only slightly from these, and will not be reported. Two-tailed tests were used throughout.

* $p < .05$.

** $p < .01$.

METHOD

Subjects

The data were collected as part of a longitudinal study of organizational change in a Southern oil refinery.[3] The maintenance division of this refinery includes a department that differs from the others in that its employees, labeled "plannables," constitute a manpower pool with no permanent work group structure. The work done by this department may be in any part of the large plant, and is scheduled in advance for brief periods only. Men are assigned to jobs on an ad hoc basis, with frequent job changes according to refinery needs. They may spend only a few days or as much as several months on any one job assignment. A plannable does not work as a member of a stable work group; he works with whoever else happens to be assigned with him to a particular job. A given plannable works under many different supervisors, depending on what job he is given. In contrast to the fluid situation of the plannables, the "assigned" wage-earners have permanently assigned posts in their departments, work in fairly stable groups, and have relatively enduring relationships with their supervisors. Data collected from 151 plannables in one department and 582 assigned employees in two departments were used in the present analysis.

Measures

Machine scorable paper-and-pencil questionnaires were administered on company time in the Spring of 1968. The questionnaire included 112 5-point Likert scales. It measured the following classes of variables:

Causal variables. These include independent variables that can be altered by management and that determine the course of developments within an organization and the results it achieves. The causal variables measured included leadership exercised by both supervisors and peers, company concern for employee welfare, general administration, and technology. The dimensions of leadership measured were support, goal emphasis, work facilitation, and interaction facilitation (Bowers & Seashore, 1967).

Group-intervening variables. These reflect the quality of behavior within work groups. The items measuring these variables have been derived from a conceptual framework found useful in previous research and utilize the reference phrase "your work group." Aspects of work group functioning measured included coordination, decision making, communication, control, cohesiveness, trust, and perceived work group effectiveness.

Contextual variables. These differ from group-intervening variables in that they refer to behavior that extends beyond the immediate work group and involves such aspects of organizational functioning as relationships between departments and events cutting across more than one hierarchical level within the organization. These variables can be thought of as the accumulated effects of the ways in which other groups in the organization function, and as a context for the behavior of the focal work group.

---

[3] The larger study is a multicompany research program undertaken by the Business and Industry Group of the Center for Research on the Utilization of Scientific Knowledge, Institute for Social Research, University of Michigan.

End-result variables. These are dependent variables which reflect the achievements of the organization. Productivity and other performance data were not available for the present analysis, but questionnaire measures of employee satisfaction and mental health provide some indication of results being achieved by the organization in terms of psychological return to its members.

Likert (1967) has elaborated the nature of causal, intervening, and end-result variables, and their interrelationships across time. The contextual variables described above were categorized separately for the present analysis after examination of the data indicated that this would be useful. Thus, Likert's "intervening variables" have been divided into two categories labeled contextual and group-intervening.

RESULTS

Causal Variables

Thirteen items measure the four supervisory leadership factors. The only two differences between plannables and assigneds that are significant indicate that assigned supervisors show their men how to improve their performance to a greater extent ($p < .01$), and plannable supervisors hold group meetings more often ($p < .01$). Each employee was asked what he thought his immediate supervisor needed in order to be a better manager; the only one of seven such items that yields a significant difference indicates that the plannables more than the assigneds felt that their supervisors have a situation that lets them do what they already know how to do and want to do. None of the nine additional questions regarding administration, company technology, technical skill of supervisors and peers, or company concern for employee welfare yields a significant difference.

Table 1 shows that the assigneds are higher on all 11 items measuring peer leadership and significantly higher on 5 of them.

Contextual Variables

Table 2 shows that the plannables score higher on 11 of the 12 measures of organizational context. Furthermore, one measure of each contextual variable reveals a statistically significant difference.

Figure 1 contrasts the distribution of control reported by the two groups. The plannables attribute significantly less influence to the first level of supervision, and significantly greater influence to top management.

TABLE 2

MEAN SCORES FOR PLANNABLES AND ASSIGNEDS ON MEASURES OF CONTEXTUAL VARIABLES

Contextual variable	Work group structure				t
	Plannable		Assigned		
	$\bar{X}$	SD	$\bar{X}$	SD	
Coordination					
Departments plan, coordinate well	3.19	1.01	3.06	1.10	
Handle department problems well	3.31	.86	3.14	1.00	
How department problems handled	3.90	1.14	3.51	1.31	3.13***
Communication					
You get information from other units	2.78	1.08	2.50	1.13	2.59***
Superiors open to ideas	3.06	1.16	2.92	1.12	
Motivation					
Conditions encourage hard work	3.34	1.03	3.10	1.16	2.19*
Participation in decision making					
Objectives set jointly	2.57	1.35	2.71	1.44	
Decision levels optimum	3.20	.96	3.05	1.06	
Decision makers seek ideas	2.55	1.07	2.45	1.10	
Decision makers get information	2.90	1.00	2.75	1.15	
Lower level problems known	2.83	1.16	2.58	1.18	2.28**
Conflict handling					
Disagreements worked through	3.53	.97	3.31	1.08	2.19*

* $p < .05$.
** $p < .025$.
*** $p < .01$.

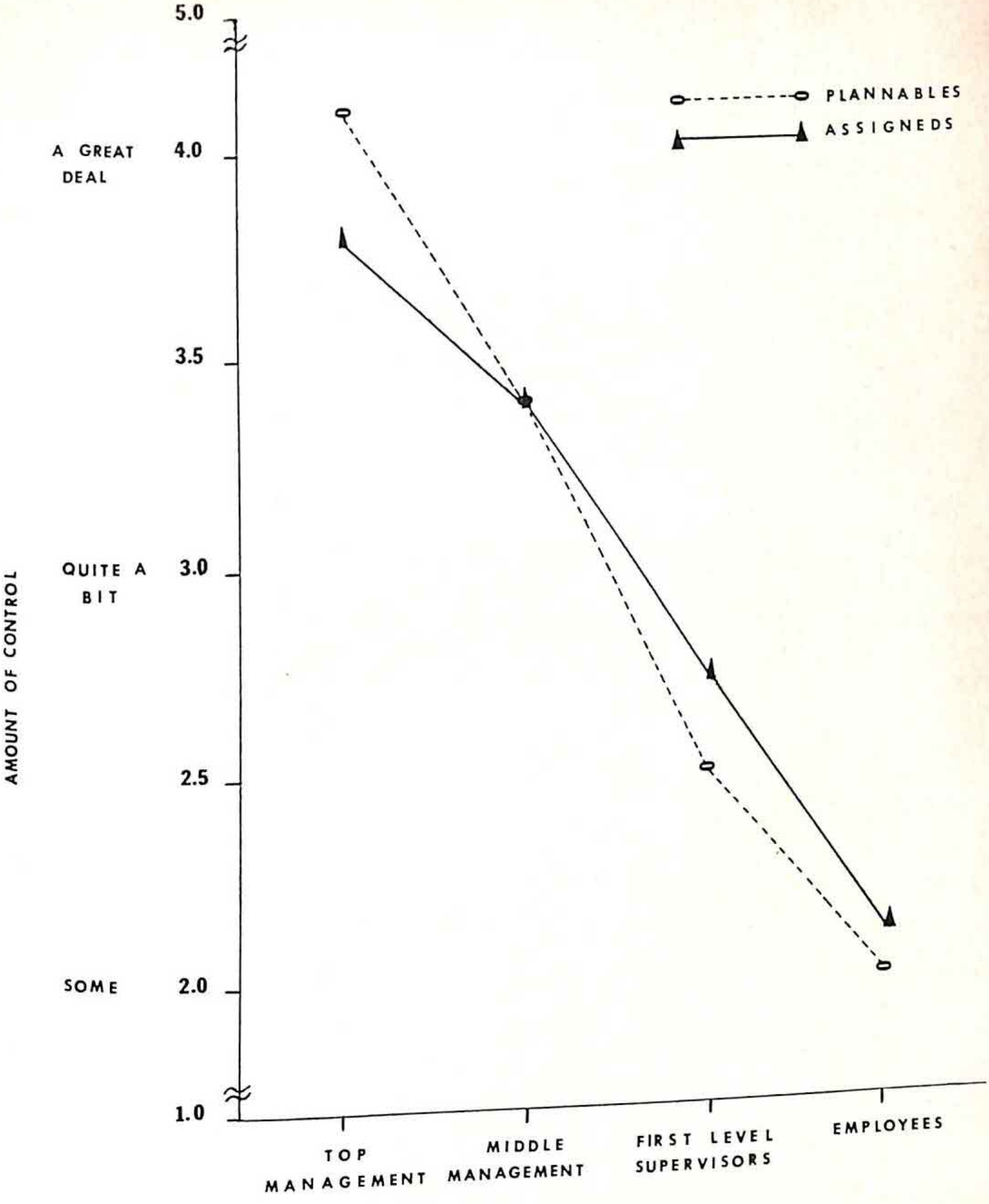


FIG. 1. Control curves for plannables and assigneds. (Significance levels of differences between plannables and assigneds in amount of control attributed to different hierarchical levels: top management, $p < .05$; first-level supervisors, $p < .025$; differences at other levels are not significant.)

Group-Intervening and End-Result Variables

Single-item measures of seven aspects of work group functioning detect no differences between plannables and assigneds. Similarly, there are no significant differences on measures of satisfaction with job, supervisor, work group, company, and pay, nor on two items measuring mental health.

DISCUSSION

The plannables and assigneds do not differ in any consistent way in supervisory leadership. Hence, this causal variable could not have exerted differential influence on other variables measured for the two groups. However, stability of group membership is a causal variable that can explain all of the differences obtained. Thus, the higher levels of assigned peer leadership would be expected, since assigned work group membership is sufficiently stable to enable peer leadership to emerge. The plannables' unstable group structure handicaps the development of peer leadership. The greater frequency of meetings held by plannable supervisors might be a response to this situation. Since the temporary and changing nature of plannable groupings does not allow for the development of a group tradition, or automatic performance of role behavior, frequent meetings would seem much less dispensable there than in the assigned situation where changes occur less frequently and past ways of doing things more readily transfer from one time period to the next. On the other hand, contextual variables, being aspects of relations among work groups throughout the organization, would be strengthened by constant movement about the refinery and more frequent contact with members of other work groups and departments, which characterize the plannables. The fluidity of the plannable situation can also explain why plannables report less need for a freer situation for their supervisors, since plannable supervisors, like their subordinates, are a pool of floaters without clear-cut work group structure. Finally, both the lower level of control exercised by plannable supervisors and the reduced extent to which they show their subordinates how to improve their performance could be due to the constant flux of workers in the plannable pool, making it a slippery situation in which it is difficult for first-line supervisors to practice close supervision.

With regard to the group-intervening variables, Likert's model fails to explain the findings. Although the assigneds have stable work group membership and report more peer leadership, they do not surpass the plannables in group-intervening variables. Apparently, the advantages accruing to the assigneds from their higher peer leadership are offset by the advantages enjoyed by the plannables in terms of a more favorable context. The net result is similarity between assigneds and plannables in measures of work group functioning and end-result variables. This interpretation assumes that context and peer leadership operate in combination to influence group-intervening variables. The manner in which leadership and context interact is not clear from the data available. One could argue that the plannables, due to their more favorable context, have managed to score as high as the assigneds on group-intervening and end-result variables despite the lower level of plannable peer leadership. This interpretation could be defended on grounds of a linkage mechanism. Likert calls individuals who are members of more than one work group "linking-pins." Conceiving of nonsupervisory personnel as potential link-pins makes possible a great expansion of the total amount of linkage within the organization. In an extreme case such as the plannable department, every member could be a link-pin. The higher context scores for plannables suggest the hypothesis that unstable work group structure induces lateral link-pin behavior on the part of nonsupervisory employees, enhancing their total system perspective and raising their context scores, thereby offsetting the costs of the loss in peer leadership that results from work group instability.

The higher context scores of the plannables and the similarity of the plannables and assigneds in measures of group-intervening and end-result variables are evidence that unstable work group membership does not necessarily impair a department's functioning. In their temporary groupings the plannables have evidently "learned to bear the absence of more enduring work relationships [Bennis, 1968, p. 75]" without major disruption of organizational processes. These findings thus suggest the viability, though not the superiority, of organic-adaptive structure as described by Bennis.

REFERENCES

BENNIS, W. G. *Changing organizations*. New York: McGraw-Hill, 1966.

BENNIS, W. G. Beyond bureaucracy. In W. G. Bennis & P. E. Slater (Eds.), *The temporary society*. New York: Harper & Row, 1968.

BOWERS, D. G., & SEASHORE, S. E. Peer leadership within work groups. *Personnel Administration*, 1967, **30**, 45–50.

LIKERT, R. *New patterns of management*. New York: McGraw-Hill, 1961.

LIKERT, R. *The human organization: Its management and value*. New York: McGraw-Hill, 1967.

(Received February 2, 1970)

JOURNAL OF APPLIED PSYCHOLOGY

JUNE 1971 VOL. 55, NO. 3

Contents

ARTICLES

Effects of Job Enlargement and Job Change on Contiguous but Nonmanipulated Jobs as a Function of Workers' Status: *Ronald C. Bishop and James W. Hill* 175

Long-Term Impact of Employee Participation in the Development of Pay Incentive Plans: A Field Experiment Revisited: *Kenneth C. Scheflen, Edward E. Lawler III, and J. Richard Hackman* 182

Testing Portions of the Porter and Lawler Model Regarding the Motivational Role of Pay: *Jay R. Schuster, Barbara Clark, and Miles Rogers* 187

Contingency Model of Leadership Effectiveness: Some Experimental Results: *George Graen, James Burdeane Orris, and Kenneth M. Alvares* 196

Note on the Methodology of the Graen, Orris, and Alvares Studies Testing the Contingency Model: *Fred E. Fiedler* 202

Contingency Model of Leadership Effectiveness: Some Methodological Issues: *George Graen, James Burdeane Orris, and Kenneth M. Alvares* 205

Regression Analysis of Three Variations of the Two-Factor Theory of Job Satisfaction: *Charles L. Hulin and L. K. Waters* 211

Expectancies as Determinants of Performance: *Abraham K. Korman* 218

Development of a Measure of Unemployability among Restored Psychiatric Patients: *Donald J. Searls, Lowell T. Wilson, and R. W. Miskimins* 223

Sources of Referral and Employee Turnover: *Martin J. Gannon* 226

Personnel Selection and the Prediction of Error: *I. van W. Raubenheimer and Joseph Tiffin* 229

Comparison of "Joint Moderators" in Three Prediction Techniques: *Sheldon Zedeck, C. J. Cranny, Carol A. Vale, and Patricia Cain Smith* 234

Communicating Directional Information with an Auditory Display: *J. Richard Simon and John L. Craft* 241

Linguistic Description of Auditory Signals: *Patricia Wright* 244

Binocular Coordination in Reading: *Karl U. Smith, Robert Schremser, and Vernon Putz* 251

MONOGRAPH

Employee Reactions to Job Characteristics: *J. Richard Hackman and Edward E. Lawler III* 259

LIST OF MANUSCRIPTS ACCEPTED 217

© 1971 by the American Psychological Association

No. 351
26.8.71
Lu
Bureau Ed. P

INFORMATION FOR CONTRIBUTORS

Style. A professional article should possess certain characteristics: (*a*) conciseness and an apparent respect for reader time; (*b*) unambiguous and simple vocabulary with technical and erudite words used only when simpler ones would obviously be inadequate; (*c*) conformity to accepted technical style in tables, terminology, and references; (*d*) conclusions that are clearly related to the evidence presented. The reader should be led, step by step, from a statement of problem or purpose, through analysis of evidence, to conclusions and implications. Authors are encouraged to consult the little book entitled *The Elements of Style* by W. Strunk, Jr. and E. B. White (New York, Macmillan, 1962).

Criteria for evaluation. Manuscripts will be evaluated on the basis of several criteria, including: (*a*) significance in contributing new knowledge to the field, (*b*) technical adequacy, (*c*) appropriateness for the *Journal of Applied Psychology*, and (*d*) clarity of presentation.

Format. Manuscripts must be prepared in the format described in the *Publication Manual of the American Psychological Association* (1967 Revision), obtainable for $1.50 from the American Psychological Association, 1200 17th Street, N.W., Washington, D. C. 20036. Articles not prepared in this manner cannot be reviewed. Special attention should be given to the section on typing the manuscript (p. 48) and to the sections on tables, figures, and references. Note that *all copy must be double spaced*, including references, title, figure captions, etc. The senior author's last name should appear on each page in the upper left-hand corner. In reference lists, give journal titles in full; do not abbreviate.

Copies. All manuscripts must be submitted in triplicate. One copy should be the original typed copy, the other two, clear carbon copies or photo reproductions, only. Authors should check the final typing carefully and retain a copy of the manuscript as a precaution against loss in the mail.

Length. Before preparing a manuscript, the author should check several recent issues to get an idea of the approximate length of regular articles published in the *Journal of Applied Psychology*. (One printed page equals roughly three double-spaced typewritten manuscript pages.) A few longer articles of special significance may be printed from time to time as monographs. Occasionally the Journal will have a section of "Short Notes" featuring brief reports on studies which make some methodological contribution or constitute an important replication.

Abstracts. Each copy of the manuscript must be accompanied by an abstract of 100–120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or they may be obtained from the Editor or from the APA Central Office.

Figures. The original drawing of a figure, or an 8 × 10-in. glossy print of the drawing, is needed for publication and must be submitted with the original typed copy of the manuscript. Duplicate copies of the figure may be photographic or pencil-drawn copies. Figures must be hand-lettered; typewritten lettering is not acceptable.

Early publication. Regular articles are published as nearly as possible in the order in which they were received. Although everything possible is done to keep publication lag to a minimum, at present there is a 10–12-mo. period between the time of receipt and publication of an article. The author of an accepted manuscript may secure early publication, usually in the first issue to go to press after its acceptance, by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication will be sent with the letter of acceptance.

Reprints. No gratis reprints are supplied. Reprints may be ordered from the printer prior to publication.

National Auxiliary Publications Service (NAPS). To save journal space, while at the same time making all necessary materials available to those with a special interest in an article, supplementary materials (such as extensive tables or copies of instruments) may be deposited with the NAPS. Copies of the material to be deposited should be submitted to the Editor with the manuscript and its availability indicated with the following footnote in the article:

> [3] *A five-page table giving data for each of the subgroups* (or other described material) has been deposited with the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

Journal of Applied Psychology
1971, Vol. 55, No. 3, 175–181

Diary No..........
Date..........
Fu[illegible] Py R[illegible]

EFFECTS OF JOB ENLARGEMENT AND JOB CHANGE ON CONTIGUOUS BUT NONMANIPULATED JOBS AS A FUNCTION OF WORKERS' STATUS[1]

RONALD C. BISHOP[2] AND JAMES W. HILL[3]

Southern Illinois University

The jobs of low-status workers were either enlarged or changed without enlargement in the presence of high-status co-workers whose jobs remained the same, and vice versa. Job enlargement was found generally to be of no greater influence than job change without enlargement so far as workers' job satisfaction and tension were concerned. Low-status workers tended to be favorably affected by job manipulation but responded unfavorably when their jobs were not manipulated. These opposing directions were attributed to a double Hawthorne effect. It was concluded that the effect of job enlargement in an organizational context involves a more complex combination of factors than has been previously considered. The utility of the *S* population for field studies of worker behavior was discussed.

Some serious questions concerning the effects of job enlargement have been raised by recent reviews of the literature on the subject (Hulin & Blood, 1968; Lawler, 1969). Two such questions provided the impetus for this study. Specifically, it has been suggested (Hulin & Blood, 1968) that a lack of controls in a number of job enlargement studies leave the results open to alternative interpretations. For example, Rice (1953) reports a study involving jobs in an Indian weaving mill. Initially, job assignments were highly specific and functionally assigned. Enlargement involved assignment of jobs on an area basis with workers performing more of the functions necessary in their assigned locations, instead of simply those jobs involved in a specific function as before. Increases in efficiency and decreases in damage were reported. The implications of the study were that the changes in the direction of job enlargement encouraged small group development which resulted in increased motivation.

It is possible, of course, that any job change in this situation might have produced similar results as occurred in the classic Hawthorne studies (Roethlisberger & Dickson, 1939). A third explanation might be that the job method was more efficient. Rice (1953) recognized some of these problems but was unable to control or measure the relative effects. It is recognized that field studies of this kind, for practical reasons, are unable to control many important variables, but this realization simply places an added burden on the applied psychologist to be constantly alert for situations where more controlled studies can be done within real work situations.

Another aspect of job enlargement which has received little systematic attention concerns its influence on workers who have contiguous but unenlarged jobs. Job enlargement or change may have an overall positive effect on those receiving the manipulation, but it may turn out to be dysfunctional for workers observing but not receiving the treatment. At least two studies imply that these dysfunctional effects may be a function of status levels of those whose jobs are not directly affected (Alderfer, 1967). The importance of individual status in a work group and its effect with respect to job satisfaction has been reported elsewhere (e.g., Homans, 1953; Roethlisberger & Dickson, 1939; and Vroom, 1964). Because of the tendencey of work group members to differentiate among themselves on the basis of status, this factor would be expected to have an important mediating effect on the results of any change which was introduced in an organization. The effect on the organization may in fact be to decrease rather than increase total worker performance and satisfaction. The seriousness of such effects was pointed out in a study by Bavelas, reported by Strauss (1955) where the positive effects of job enlargement were nullified by increased complaints of workers on unenlarged jobs. The final result was that the job enlargement was cancelled, and within a month, six of eight girls who had been on the enlarged task quit. Strauss, referring to the Bavelas results, concludes that "The factory is a social system made up of mutually dependent parts of a system. It may be dangerous for management to try a new approach in one small part unless it is prepared to extend this approach to the whole organization [p. 96]."

---

[1] This study was carried out in connection with a dissertation submitted by the senior author to the Department of Psychology, Southern Illinois University. The authors gratefully acknowledge the helpful suggestions of James G. Hunt and William M. Vicars.

[2] Requests for reprints should be sent to Ronald C. Bishop, Department of Management, Southern Illinois University, Carbondale, Illinois 62901.

[3] Now at Western Michigan University.

The purposes of the present study were (*a*) to compare the effects of job enlargement with those of a simple job change in a controlled field setting, and (*b*) to compare the effects of job enlargement and job change on contiguous but nonmanipulated jobs as a function of workers' status in the work group.

Job enlargement as it is used here follows the definition provided by Kilbridge (1960): "The expansion of job content to include a wider variety of tasks and to increase the worker's freedom of pace, responsibility for checking quality, and discretion for method [p. 357]." This definition is similar to what Lawler (1969) described as vertical job enlargement and what Herzberg (1968) termed vertical job loading. These definitions each emphasize the personal involvement of the worker in the successful completion of his task rather than simply giving him more tasks to complete.

Job change is here defined as a change in job content which requires a different task but no increase in responsibility or in the number of tasks performed.

Dependent variables were job satisfaction as measured by the work subscale of the Job Descriptive Index (Smith, 1967), job performance as measured by quantity and quality of output, and tension as measured by the Anxiety Differential (Alexander & Husek, 1962). Measures were taken before and after manipulation for workers whose jobs were altered as well as for workers whose jobs remained the same. A measure of tension was included among the dependent variables as a means of identifying a possible underlying correlate of changes in job satisfaction and performance. In addition to these three variables, measures of individual status in the work group were taken sociometrically before and after the job manipulation. Status was also an independent variable in that the introduction of job change or job enlargement was dependent upon the workers' status in the group.

METHOD

Research Setting

This study was conducted in the Employment Training Center at Southern Illinois University, which is a sheltered workshop for the rehabilitation of mentally and physically handicapped persons. In keeping with this purpose, clients work on jobs contracted with local industry. It is emphasized that these are real jobs and that productivity and quality standards must be maintained. Clients work from 8:00–3:30 daily with 15-min. breaks at 10:00 A.M. and 2:00 P.M. and a half-hour lunch period. Some clients work under an incentive pay system and others on an hourly basis, but during the experiment all participants were paid by the hour. Records are kept of attendance, tardiness, and grooming, as well as quality and quantity of performance. The clients have regular supervisors and are expected to clean up and maintain their work areas. One major difference between this work environment and others is that the clients often participate in university training and research projects. They are thus used to being questioned about themselves and their attitudes and do not regard research of this kind as being unusual.

Subjects

Forty-eight client-worker *S*s were randomly assigned to one of eight work groups. The task was one regularly performed by workshop clients for a local manufacturer, and each *S* had previously been assigned to a similar operation one or more times during his work experience in this setting. There were 30 men and 18 women ranging in age from 17 to 60. Length of time in attending the workshop ranged from 2 mo. to 9 yr. with average experience about 2 yr. No *S* with an IQ below 75 was included in the group. Eight *S*s had physical impairments, but none were considered serious enough to influence their performance on the assigned tasks. Twenty-six *S*s were illiterate.

Procedure

Each group worked a total of 5 days on the experimental task which consisted of sorting either nuts or bolts. Each *S* independently sorted a different size nut or bolt, but the whole group sorted either nuts or bolts. The number of correct items sorted was the measure of quantity of production, while those items incorrectly sorted or omitted were recorded as errors and served as the measure of quality of production.

On the afternoon of the second work day, each *S* was interviewed privately and asked to rate the members of his work group including himself on the dimensions of ability, intelligence, and influence. The total rating on each of these factors by each member of the work group constituted a worker's status rating within the experimental work group. On the morning of the third work day, an attitude and a tension measure were administered to each *S*: the work subscale of the Job Descriptive Index (JDI; Smith, 1967) was used to measure job satisfaction, and the Anxiety Differential (AD; Alexander & Husek, 1962) was employed as the measure of tension. Administration to *S*s whose literacy level made it necessary was conducted orally and individually. Test-retest reliability (7–10 day interval) as established on a sample from this population prior to the formation of the experimental groups was .95 for the JDI and .88 for the AD.

Midway through the third work day, either a job change or job enlargement was introduced to three members of the work group selected according to their relative status (high or low) in the group while the remaining three members continued on their originally assigned tasks. The job change consisted merely of assigning a particular size nut to *S*s who had been sorting bolts or vice versa. The job enlargement involved the additional assignment of sorting a specified nut or bolt, assembling the nut and bolt, and placing the assembly in a packing box with the admonition, "Be sure to get these right because they will be shipped out at the end of the day."

Because the enlarged task involved a more complex and longer methodological sequence than the unenlarged task, units produced under each condition were not comparable. In order to determine an adjustment factor to apply to this difference, a time study was made of both the enlarged and unenlarged versions of the experimental task and a standard time computed for performance on each. Deviations from this expected difference in standard times was used as the measure of change in quantitative performance. Also, in this way, gross differences in performance due to method were controlled.

Status as well as job satisfaction and anxiety measures were readministered on the afternoon of the fourth and the morning of the fifth work days.

At no time during the procedure were *S*s informed as to the experimental nature of the work groups, nor was there any indication that they regarded this job as being different from other work they had done before.

Design

Eight experimental groups composed of six workers each were used in the study, with four groups assigned to the job enlargement condition and four groups assigned to the job-change-without-enlargement condition. Under each job manipulation condition (job enlargement and job change), the jobs of the three high-status members were manipulated while the jobs of the three low-status members remained the same. In each of the remaining four work groups, the three low-status members' jobs were manipulated while the jobs of the three high-status members remained unchanged. This design made possible a test of the effects of job enlargement compared to job change without enlargement as a function of the workers' relative status in their work group. This was possible not only for workers whose jobs were manipulated, but for contiguous nonmanipulated workers as well.

Analyses were done on difference scores for each group. Differences between premanipulation and postmanipulation measures of quality and quantity of performance, relative status, job satisfaction, and tension were analyzed by means of a $2 \times 2 \times 2$ factorial analysis of variance.

RESULTS

Performance

While changes in quantity of output were not found to be significant under any of the experimental conditions, several significant changes in quality did occur. Table 1 shows a significant ($p < .05$) main affect for manipulated versus nonmanipulated workers. It can be seen in Table 2 that workers whose jobs were manipulated accounted for significantly more errors than their co-workers who did not undergo either job enlargement or job change. Two significant ($p < .05$) two-way interactions were obtained. The first was between the job enlargement and change conditions and whether or not the workers had their jobs manipulated. It can be seen in Table 2 that workers whose jobs were enlarged suffered a decrement in quality of performance ($t = 3.76$, $df = 6$, $p < .01$) while those whose jobs were merely changed without being enlarged showed less change. Since this difference could have been due to the greater complexity of the enlarged task, a correlation between error frequency and the levels of task complexity was computed ($r_{\text{bis}} = .37$, $p < .05$). This would suggest that error frequency was at least a partial function of a greater comlexity of the enlarged task over the unenlarged task. Nonmanipulated workers, on the other hand, showed an improvement in quality of performance under enlargement and essentially no change in performance under job change.

The apparently facilitative effect on quality of performance for nonmanipulated workers when job enlargement was introduced for their higher status co-workers was not significant. However, in all four experimental conditions, low-status workers were more re-

TABLE 1

ANALYSIS OF VARIANCE OF CHANGES IN QUALITY: ERRORS IN PERFORMANCE

Source	*df*	*MS*	*F*
Task (job enlargement-job change) (A)	1	1773.50	4.14
Level of involvement (manipulated-nonmanipulated) (B)	1	4463.17	10.41*
Worker status level (high-low) (C)	1	1844.73	4.30
A × B	1	3577.59	8.34*
A × C	1	367.76	.86
B × C	1	4069.55	9.49*
A × B × C	1	2600.69	6.07*
Error	8	428.75	

* $p = .05$.

TABLE 3

ANALYSIS OF VARIANCE OF CHANGES IN SATISFACTION SCORES: JOB DESCRIPTIVE INDEX WORK SUBSCALE

Source	*df*	*MS*	*F*
Task (job enlargement-job change) (A)	1	6.25	1.00
Level of involvement (manipulated-nonmanipulated) (B)	1	64.00	10.24*
Worker status level (high-low) (C)	1	42.25	6.76*
A × B	1	42.25	6.76*
A × C	1	9.00	1.44
B × C	1	30.25	4.84
A × B × C	1	4.00	.64
Error	8	6.25	

* $p = .05$.

sponsive (see Table 2) than were high-status workers. Also, it should be noted that under three out of four conditions, responses of nonmanipulated workers at one status level were in the opposite direction to those of manipulated workers at the other status level.

The analysis of variance of the work subscale of the JDI (see Table 3) yielded two significant main effects ($p < .05$). The first indicates that nonmanipulated workers' satisfaction with their work declined while manipulated workers' satisfaction increased (except for one condition). The other shows that higher status workers displayed significantly greater declines in satisfaction with their work as measured by the JDI than their low-status co-workers.

The third significant effect ($p < .05$) for the analysis of variance of changes in JDI work subscale responses was a two-way interaction between the job enlargement-job change conditions and the manipulated-nonmanipulated conditions. It can be seen in Table 4 that workers whose jobs were changed showed an increase in satisfaction while those whose jobs were enlarged showed a small average decrease. Workers in contiguous but nonmanipulated jobs showed a greater decrease in satisfaction under the job change condition than under the job enlargement condition.

Changes in AD responses were found to reflect less tension for those whose jobs were manipulated while reflecting more tension for workers whose jobs remained the same (see Table 5). This finding may reflect a double Hawthorne effect (i.e., facilitating and inhibiting effects) due to the recognition of those who were singled out for the job manipulation and the disappointment of those who were excluded.

The analysis of variance of changes in status ratings (see Table 6) yielded two significant ($p < .001$) main effects. Manipulated workers experienced an increase in status compared to

TABLE 2

TABLE OF MEAN DIFFERENCES IN GROUP QUALITY OF PERFORMANCE: FREQUENCY OF ERRORS

Status	Job enlargement		Job change	
	Manipulated	Nonmanipulated	Manipulated	Nonmanipulated
High	10.98	5.08	−4.89	−1.99
Low	99.45	−21.26	13.40	3.50

TABLE 4

TABLE OF MEAN DIFFERENCES IN SATISFACTION SCORES: JOB DESCRIPTIVE INDEX WORK SUBSCALE

Status	Job enlargement		Job change	
	Manipulated	Nonmanipulated	Manipulated	Nonmanipulated
High	−4.5	−3.5	.5	−3.0
Low	2.0	−.5	6.0	−5.0

TABLE 5

ANALYSIS OF VARIANCE OF CHANGES IN TENSION: ANXIETY DIFFERENTIAL SCORES

Source	df	MS	F
Task (job enlargement-job change) (A)	1	225.00	.68
Level of involvement (manipulated-nonmanipulated) (B)	1	2756.25	8.29*
Worker status level (high-low) (C)	1	756.25	2.27
A × B	1	72.25	.22
A × C	1	1056.25	3.17
B × C	1	169.00	.51
A × B × C	1	576.00	1.73
Error	8	332.625	

* $p = .05$.

TABLE 6

ANALYSIS OF VARIANCE OF CHANGES IN STATUS IN GROUP

Source	df	MS	F
Task (job enlargement-job change) (A)	1	.77	.11
Level of involvement (manipulated-nonmanipulated) (B)	1	606.39	84.92*
Worker status level (high-low) (C)	1	319.52	44.75*
A × B	1	19.14	2.68
A × C	1	23.97	3.33
B × C	1	1.89	.26
A × B × C	1	1.89	.26
Error	8	7.14	

* $p = .001$.

nonmanipulated workers whose status declined. When status changes occurred following the introduction of the job manipulation, lower status workers were more likely to show upward changes while higher status workers declined. This result can be largely attributed to the measurement technique employed which made it difficult for higher status workers to improve their positions in the group as well as for lower status workers to decline in status.

In summary, results showed quality rather than quantity of performance to be affected by job enlargement. This finding is similar to those of Kilbridge (1960), Kuriloff (1966), and Biganne and Stewart (1963). However, in the present study, job enlargement generally was found to be of no greater influence on dependent variables than the simple job change. It should be pointed out too that contiguous but nonmanipulated workers often showed changes in the opposite direction from those whose work was redesigned. In addition, group status was an important factor with low-status workers found to be more responsive to change than high-status workers in most cases. The status level of the workers whose jobs were manipulated was a significant factor influencing changes in quality, satisfaction with work, anxiety, and rank in the work group.

DISCUSSION

The similarity between job change and job enlargement effects on most dependent variables in this study raises an important question: How often might the effects attributed to job enlargement in other studies be due instead simply to a job change? Perhaps merely selecting a group for a new task accounts for changes in satisfaction, tension, and status like those found in this study. This might be another example of the Hawthorne effect where special attention is seen as a source of recognition for those selected, and not being selected is cause for disappointment (negative Hawthorne effect). Of course, activation theory (Scott, 1966) would predict similar results. However, some difficulty would be encountered in explaining why a more complex change (i.e. job enlargement) did not have a greater effect than a simple job change.

The results concerning manipulated versus contiguous but nonmanipulated work groups are a function of the status level of the workers. In general, however, nonmanipulated but contiguous workers showed decreases in status and satisfaction with work and increases in anxiety. The results support those of Strauss (1955) and Alderfer (1967) concerning the negative effects of changes in job design on contiguous but nonmanipulated groups of workers. One might speculate that if these results were widespread in the organization, the total effect could be dysfunctional.

The data on quality seem confounded with task complexity and make interpretation questionable. This is illustrated by the dramatic decrement in quality that occurred under the job enlargement condition and the fact that the two levels of task complexity (job enlargement and job change) were found to show a relationship to error frequency.

Why high-status workers generally should show a decline in satisfaction is not clear. Perhaps they were only too well aware of their own high status and were threatened by any job change in which they were directly involved or merely watching. Both Vroom (1964), Homans (1953), and others have indicated that high-status workers are more satisfied with their jobs than low-status workers, and this might make them more resistant to any change than low-status workers. Conceivably, low-status workers may perceive a direct change as an opportunity to improve their position in the group.

Anxiety decreased as status increased for those whose jobs were manipulated and the opposite was true for those whose jobs were not manipulated. In addition, nonmanipulated workers displayed declines in satisfaction while manipulated workers showed an increase overall. This suggests that tension and status may be moderators of the relationship between job satisfaction and job design. It further suggests that the so-called Hawthorne effect may be in fact a function of a perceived increase in status accompanied by a decrease in tension. These results add strong support to the findings of Vroom (1964) and Homans (1953) that status and satisfaction are positively related. In the present study, perceived increases in status were associated with increased satisfaction.

Recognizing the interaction in the present study, it was found that low-status workers were more responsive on most dependent variables to changes in their work environment than high-status workers. Fenchel, Monderer, and Hartley (1951) have reported a similar finding that individuals exhibited greater striving in groups where their status was low than in groups where their status was high. The results appear to indicate that, at least under these conditions, status must be considered when the relationship between job design and job satisfaction is examined.

The generalizability of experimental results from an *S* population such as the one used in the present study remains open to question. Certainly, this population is free of many of the constraints surrounding more normal work groups, such as high production quotas and union restrictions. Further examination of its usefulness for controlled studies of worker behavior seems warranted. Industrial research in a sheltered workshop environment can have several advantages. (*a*) The population is worth considerable interest in its own right. (*b*) The situation can be controlled by the investigator more than is usual in normal work situations. (*c*) It is a more realistic work environment than we find in experimental studies using college students and artificial work tasks. Further, as the present results indicate, these workers seem to respond to many of the same kinds of factors that other work groups do.

All of these results suggest that the effects of job enlargement on organizational performance involve a much more complex combination of factors than has been previously examined. While the number of replications under each of the experimental conditions in the present study is a bare minimum, the results would, nevertheless, suggest that the factors examined are worthy of consideration in future investigations of job enlargement.

REFERENCES

ALDERFER, C. P. The organization syndrome. *Administrative Science Quarterly*, 1967, **12**, 440–460.

ALEXANDER, S., & HUSEK, T. R. The Anxiety Differential: Initial steps in the development of a measure of situational anxiety. *Educational and Psychological Measurement*, 1962, **22**, 325–348.

BIGANNE, J. F., & STEWART, P. A. *Job enlargement: A case study.* (Research Serial No. 25) Bureau of Labor and Management, State University of Iowa, 1963.

FENCHEL, G. H., MONDERER, J. H., & HARTLEY, E. L. Subjective status and the equilibration hypothesis. *Journal of Abnormal and Social Psychology*, 1951, **46**, 476–479.

HERZBERG, F. One more time: How do you motivate employees?, *Harvard Business Review*, 1968, **12**, 53–62.

HOMANS, G. C. Status among clerical workers. *Human Organization*, 1953, **12**, 5–10.

HULIN, C. L., & BLOOD, M. R. Job enlargement, individual differences, and worker responses. *Psychological Bulletin*, 1968, **69**, 41–55.

KILBRIDGE, M. D. Reduced costs through job enlargement. *Journal of Business*, 1960, **33**, 357–362.

KURILOFF, A. H. *Reality in management.* New York: McGraw-Hill, 1966.

LAWLER, E. E., III. Job design and employee motivation. *Personnel Psychology*, 1969, 22, 426–435.

RICE, A. E. Productivity and social organization in an indian weaving shed. *Human Relations*, 1953, 6, 297–329.

ROETHLISBERGER, F. J., & DICKSON, W. J. *Management and the worker—An account of a research program conducted by the Western Electric Company, Chicago.* Cambridge: Harvard University Press, 1939.

SCOTT, W. Activation theory and task design. *Organization Behavior and Human Performance*, 1966, 1, 3–26.

SMITH, P. C. The development of a method of measuring job satisfaction: The Cornell studies. In E. A. Fleishman (Ed.), *Studies in personnel and industrial psychology*. Homewood, Illinois: Dorsey Press, 1967.

STRAUSS, G. Group dynamics and intergroup relations. In W. F. Whyte (Ed.), *Money and motivation*. New York: Harper, 1955.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

(Received March 9, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 182–186

LONG–TERM IMPACT OF EMPLOYEE PARTICIPATION IN THE DEVELOPMENT OF PAY INCENTIVE PLANS:

A FIELD EXPERIMENT REVISITED[1]

KENNETH C. SCHEFLEN,[2] EDWARD E. LAWLER III,[3] AND J. RICHARD HACKMAN

Yale University

In an earlier study Lawler and Hackman examined the effects of worker participation in the development of pay incentive plans. In the original study, three work groups developed their own incentive plans to reward high attendance, and identical plans were then imposed by company management in two other work groups. A significant increase in attendance was found during the first 16 wk. following implementation of the plans only in the groups where the plans were participatively developed. Data reported in the present study cover a 12-wk. period beginning 1 yr. after the original plans had been installed. After the data reported in the earlier study had been collected, the incentive plans were discontinued by company management in two of the three participative groups. The present results show that attendance dropped below pretreatment levels in these two groups, and that attendance continued high in the third participative group. An increase in attendance was found after 1 yr. in those groups where incentive plans had been imposed by company management.

In a previous study, Lawler and Hackman (1969) examined the effects of employee participation in the development of pay incentive plans. The *S*s were part-time unskilled workers who cleaned buildings in the evening. In the experiment, three autonomous work groups developed their own incentive plans to reward good job attendance (participative condition). These plans were then imposed by the company on two other work groups (imposed condition). There were two control groups: one group talked with the researchers about job attendance problems but received no treatment, and the other received no attention from the researchers. Attendance was monitored for 16 wk. following the introduction of the new pay plans. A significant increase in attendance appeared only in the participative condition. Before the change, the average employee in the participative condition worked 88% of his scheduled hours; after the plan went into effect, the average employee worked 94% of his scheduled hours. In the imposed condition, the comparable pretreatment figure was 83% and the posttreatment figure also was 83%. No changes were observed in the control groups.

It is possible that the increase in attendance in the participative condition was due solely to an initial burst of enthusiasm on the part of the workers in the participative groups and that the effects observed during the 16-wk. postexperimental period were only temporary. Alternatively, it is possible that the imposed plans were, in the long run, equally as effective as the participative plans—but that it simply takes longer for an imposed plan to have effect. To examine these questions, follow-up attendance data were gathered from the original work groups approximately 1 yr. after the pay plans were initially implemented.

In addition, the study examines the effects of a naturally occurring reversal of the original experimental treatment. After the pay plans had been in effect for 6 mo. in the case of one group and 11 mo. in the case of another group, company management unexpectedly discontinued the incentive plans in two of the three participative groups. Workers in both of these groups had shown a significant increase in attendance when the new plans had been instituted. The present data show the effects on the attendance of the employees in these groups when they reverted to the preexperimental pay arrangements.

---

[1] This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Now at the Base Cascade Corporation.

[3] Requests for reprints should be sent to Edward E. Lawler III, Department of Administrative Sciences, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.

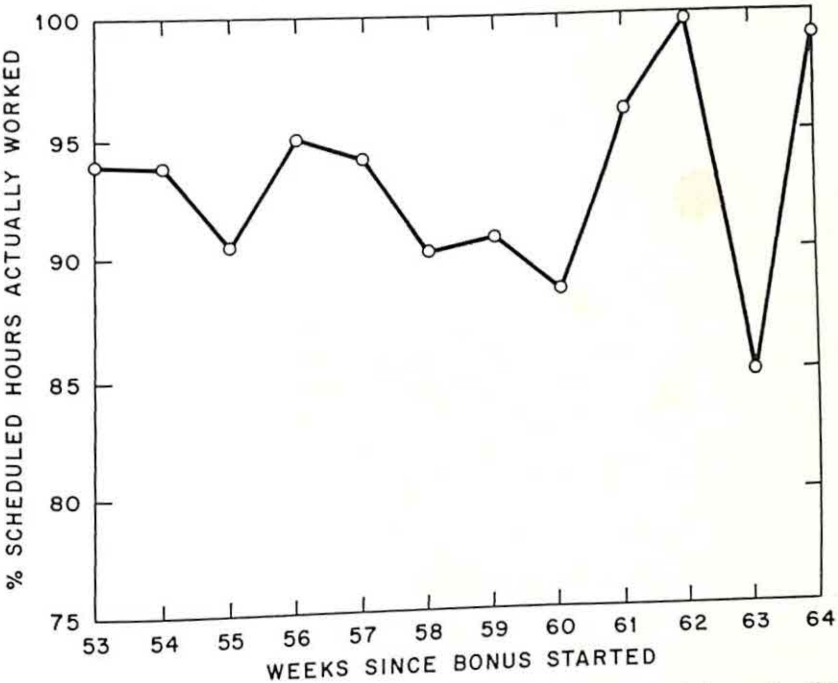


FIG. 1. Attendance of one participative group for a 12-wk. period 1 yr. after the plan started. (Attendance is expressed in terms of the percentage of the hours scheduled to be worked that were actually worked.)

METHOD

The research site was a small company that provides building maintenance services on a contract basis. The *S*s were part-time employees who cleaned buildings during the evening. The *S*s worked in groups which ranged in size from 2 to 25 and which were responsible for doing cleaning work in one building. Although the work groups did similar work, they were highly autonomous and there was very little communication among employees in different work groups.

The employees tended to have very low educational levels and most were members of minority groups. They ranged in age from 16 to over 70 yr. Although the employees from whom data were collected in the present study are very similar to those in the original study, about 60% of the original participants had left the company by the time data collection for this study had been completed. This high turnover rate is characteristic of all maintenance employees in the company and was not higher in the experimental and control groups used in this research. The research site and *S*s are discussed in greater detail by Lawler and Hackman (1969).

Attendance data were collected for the present study from the five work groups which had participated in the earlier study. Two of the groups had incentive plans imposed on them by management in the earlier study, and the plans were still in effect when the present data were collected. One group had participatively developed an incentive plan in the earlier study, and the plan was still in effect. Two groups had participatively developed incentive plans in the earlier study, but the plans subsequently had been discontinued by company management.

A weekly percentage attendance figure was obtained for each group by dividing the number of hours scheduled to be worked each week into the number of hours actually worked. In addition, employees in three of the groups were given semistructured on-the-job interviews about their reactions to the incentive plans, and company managers were interviewed regarding their views about the way the plans had been introduced and how they were functioning.

RESULTS

Results for the participative groups are presented in Figures 1 and 2. In the single participative group in which the pay plan was still in effect at the time of the present data collection, the average employee worked 93% of his scheduled hours (see Figure 1). This compares to 94% during the 16-wk. observation period immediately following the introduction of the plan, and 88% prior to the introduction of the plan. Thus, the plan appears still to be effective for this group.

Data for the two participative groups for which the plans had been discontinued by management are shown in Figure 2. (Although the plans were discontinued at different times,

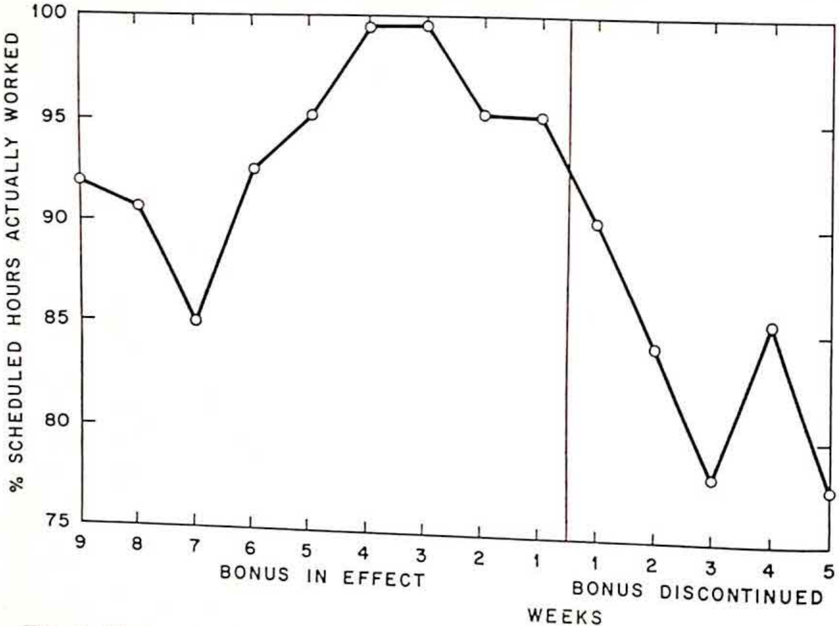


FIG. 2. Mean attendance of two participative groups for the 9 wk. before the plans were ended and for the first 5 wk. after they were ended. (Attendence is expressed in terms of the percentage of hours scheduled to be worked that were actually worked.)

the data for the two groups have been combined.) In the weeks immediately prior to discontinuation, average attendance in these two groups was 92%; in the weeks after the plans had been discontinued, attendance dropped to an average of 82%. The 92% attendance is

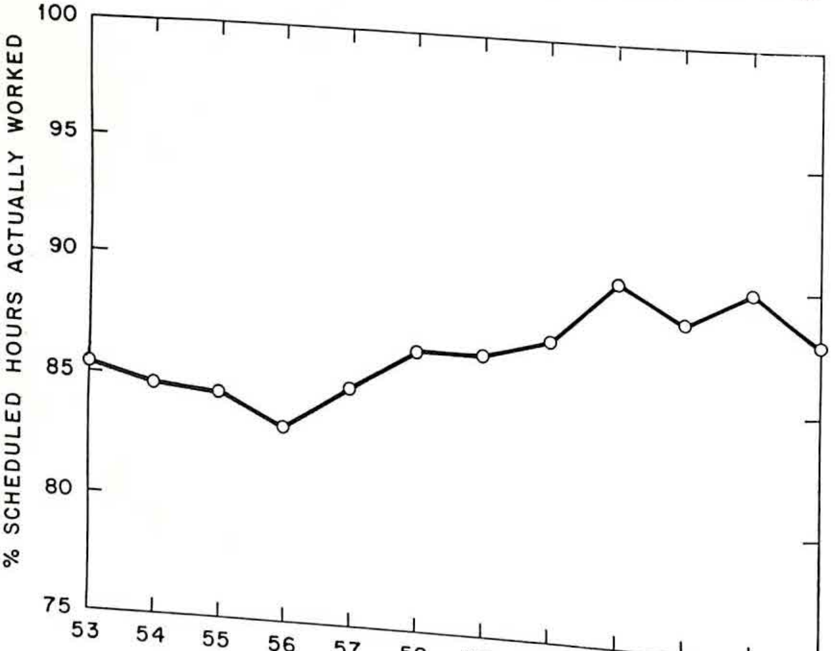


FIG. 3. Mean attendance of the two imposed groups for a 12-wk. period 1 yr. after the plans started. (Attendance is expressed in terms of the percentage of the hours to be worked that were actually worked.)

comparable to the 94% average obtained for the participative groups in the original study. The drop from 92% to 82% is statistically significant (by median test, $\chi^2 = 7.78$, $p < .01$), suggesting that discontinuation of the plan did decrease average employee attendance.

Data from the two imposed groups are presented in Figure 3. During the 12-wk. observation period, attendance in these groups averaged 87% compared to 83% for the 12-wk. period immediately preceding the installation of the plan (by median test, $\chi^2 = 6.00$, $p < .05$). Thus, attendance in the imposed groups is somewhat higher 1 yr. after the plans were imposed than it was immediately following their implementation.

DISCUSSION

It is significant that the attendance data were still high in the participative groups 1 yr. after the plans were installed, despite the fact that a majority of the workers in these groups had not themselves been involved in the development of the plans. It would appear that the norms for good attendance which seemed to have developed in the participative groups when the plans were introduced were effectively passed on to new workers as they joined the work groups. In one group, for example, employees expressed feelings of ownership of the plan despite the fact that most of the group members had not actually participated in the design of the original plan. These data suggest that the initial success of the participative plans was not simply a function of an initial and temporary enthusiasm about the opportunity to participate in the development of a pay plan, and that groups seem to have norms that do not depend on a stable membership for their maintenance.

The small increase in attendance observed for the groups in which the pay plans were imposed was not entirely unexpected. Although employees did not respond to these plans immediately after they were introduced, the company did administer the plan fairly and consistently throughout the year preceding the present data collection. Experience with the plan over this longer time period may finally have convinced the workers that the plan was on the level, and that if they came to work more regularly they would receive more money. Further, the workers in the imposed groups simply may have taken longer than their colleagues in the participative groups to learn the mechanics of how the plans worked. The net result, it appears, was that both the workers' trust of management and their understanding of the plan and its implications for the amount of pay they could earn increased slowly over time as did their attendance. It should be noted, however, that even after a year of slow improvement, the imposed groups had not reached the level of attendance attained and maintained by the participative groups.

Probably the most compelling finding of the present study is the large and immediate drop in attendance when the pay plans were discontinued in two of the three participative groups. The decision of company management to drop the two plans created a second manipulation in the study and, in effect, produced a natural field experiment. This second manipulation provides dramatic evidence for the effectiveness of the participative plans: dropping them reduced attendance just as substantially as installing them had increased it.

The decision by company management to discontinue two of the plans came as a complete surprise to the researchers. Financial considerations probably did not play a role in the decision, since hourly rates were raised when the plan was discontinued. Indeed, discontinuance of the plans appeared to cost the company more money in increased hourly wages than it saved in unpaid bonuses.

Further, the decision did not seem to hinge on misperceptions of company managers about the effectiveness of the plans. The researchers had shown the relevant managers the article written about the earlier study and most of these individuals accepted the fact that the participative plans were working as intended. Interviews with several managers responsible for administering the plans did suggest that these managers felt little or no ownership of the plans or commitment to them. They had been told by top management to work with the researchers in instituting the experimental program, but they themselves had little to do with developing it. It is not surprising, therefore, that these managers felt little commitment to the plans and that they were ready to discontinue them when the first opportunity presented itself.

Viewing the long-term impact of the experimental program on organizational effectiveness, it appears that the researchers' failure to gain the commitment of the managers was a very significant one indeed. The two imposed pay plans persisted and even showed a small increase in effectiveness 1 yr. later. But two of the three plans in which employees participated to a significant degree were dropped. Thus, the net result is that the three participative plans were not as effective over the long term as were the imposed plans. The reason, it should be emphasized, is not that employees respond better to imposed plans. The data suggest just the reverse. In summary, it appears that the present data support the view that long-lasting change can be introduced effectively in organizations only when the changes are accepted and owned by all levels—employees and managers alike—who are or will be involved in the new program.

REFERENCE

LAWLER, E. E., III, & HACKMAN, J. R. Impact of employee participation in the development of pay incentive plans: A field experiment. *Journal of Applied Psychology*, 1969, **53**, 467–471.

(Received December 22, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 187–195

TESTING PORTIONS OF THE PORTER AND LAWLER MODEL REGARDING THE MOTIVATIONAL ROLE OF PAY[1]

JAY R. SCHUSTER,[2] BARBARA CLARK, AND MILES ROGERS

System Development Corporation, Santa Monica, California

Hypotheses regarding the motivational role of pay as set forth by Lyman Porter and Edward Lawler in their book, *Managerial Attitudes and Performance*, were tested. Two key hypotheses were confirmed: (*a*) the more employees believe that performance factors influence pay, the harder they will work to improve their performance; (*b*) individuals who see pay as a satisfier will try to perform more satisfactorily. These hypotheses were confirmed in an environment where the pay-performance relationship is probably stronger than in the organizations used by Porter and Lawler in testing their hypotheses. However, this environment may account for our finding that effort functions differently than in the Porter and Lawler theoretical model.

In their book, *Managerial Attitudes and Performance*, Lyman Porter and Edward Lawler (1968) have established a unique precedent for academicians performing psychological research on pay. They have contributed to the industrial practitioner's capacity for designing an equitable reward system by testing, in a total of seven diverse private and governmental organizations, various hypotheses based upon their theoretical model of a performance-reward system. This article describes our tests of some of the hypotheses dealing with the motivational role of pay. They were made in one private company rather than in a number of diverse organizations.

The company was selected for the present study because it seemed to be attempting to pay or reward its employees on the basis of performance, with the better performers receiving the larger increases (see Table 1). The objective was to test the Porter and Lawler hypothesis through research conducted in an environment that paid for performance to a greater extent than did the environments researched by Porter and Lawler. In regard to their environments, Porter and Lawler said: "for companies in our sample, pay is being dispensed at least as much for trying hard (i.e., effort) as for accomplishment (i.e., performance) [p. 158]."

We used Porter and Lawler's terminology of variables and the interrelationships of variables that they derived from the theoretical model to test the hypotheses (see Figure 1). These variables, definitions, and interrelationships are as follows:

Value of reward: attractiveness or desirability to the individual of the potential outcome of his behavior in the work situation.

Perceived effort-reward probability: an individual's expectations concerning the likelihood that given amounts of rewards depend upon given amounts of effort on his part.

Effort: how hard the individual is trying.

Abilities and traits: the relatively stable, long-term individual characteristics that represent the individual's currently developed power to perform.

Role perceptions: the direction toward which an individual feels he should expend his effort in order to perform satisfactorily.

Performance: the result of effort modified by individual abilities and traits and role perceptions.

Rewards: positively valued intrinsic or extrinsic outcomes resulting from satisfactory performance.

Perceived equitable rewards: level or amount of rewards the individual feels he should receive as a result of a given level of performance, or which an individual feels should be attached to a particular job.

---

[1] The authors wish to thank Edward E. Lawler III of Yale University for his comments and suggestions on an earlier draft. The authors also wish to thank Raymond Enright of System Development Corporation's Research Directorate for his valuable assistance in the statistical analysis of the data.

This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Jay R. Schuster, Manager-Compensation and Personnel Planning, System Development Corporation, 2500 Colorado Avenue, Santa Monica, California 90406.

TABLE 1

ACTUAL MONTHLY SALARY INCREASES GRANTED IN THE SURVEY FIRM DURING CALENDAR YEAR 1969, BY SUPERVISORY PERFORMANCE RATING

Supervisory performance rating[a]	% respondents receiving monthly salary increases[b]			
	$50 and below	$51–$100	$101–$150	$151 and above
High	4.0	35.1	47.0	13.9
Middle	12.0	64.3	18.6	5.1
Low	48.5	43.6	5.9	2.0

[a] Performance ratings were determined by first having supervisors rank-order the employees reporting to them on the basis of performance or contribution to supervisory goals and objectives. Supervisors then met to integrate the rankings based upon cross-organizational employee contributions. Levels were assigned on the following basis: high, top 30% of performers in population contacted ($n = 202$); middle, next 50% of performers in population contacted ($n = 393$); low, bottom 20% of performers in population contacted ($n = 101$).

[b] $\chi^2 = 197.45$, $df = 6$, $p < .001$.

Satisfaction: the extent to which the rewards meet or exceed the perceived equitable reward level.

Interrelationships of variables. Testing their hypotheses regarding the nine variables in the model led Porter and Lawler to conclude that the more an individual values a reward and the more he sees his efforts as leading to this reward, the more effort he will expend. If the performer has accurate role perceptions, his expended effort will be converted into meaningful performance which will in turn obtain valued rewards for the individual. Performance also leads to satisfaction if it results in the amount of the valued rewards matching the level of rewards considered by the individual as equitable for his performance level. If the reward granted is less than the perceived equitable reward expectation level, the performer will be dissatisfied. Rewarding performance commensurately reinforces perceptions of the probability that efforts are rewarded and are thus worthwhile. However, the value of a reward to the giver and a reward's value as perceived by the recipient may differ. If there is a difference, the probability in the recipient's mind that his increased efforts result in increased rewards will lose credibility since the recipient will not see an individually valued reward-performance connection. When a giver and recipient of a reward do view it in the same way, the recipient's satisfaction with his reward will affect the value of future rewards. For example, such higher order needs as esteem, autonomy, etc., become more attractive the more an individual feels satisfied with a given reward level. The Porter and Lawler (1968) hypotheses retested in this study are listed below.

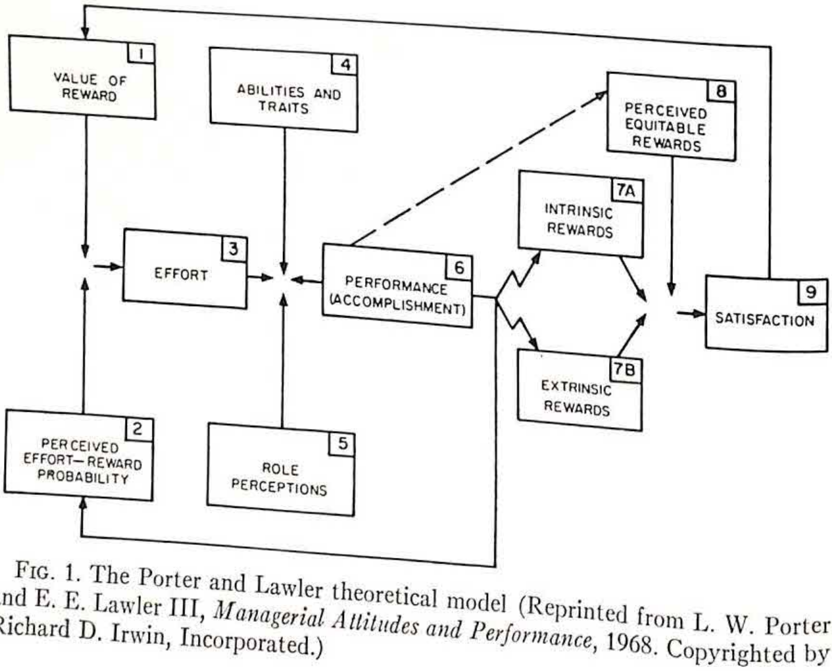


FIG. 1. The Porter and Lawler theoretical model (Reprinted from L. W. Porter and E. E. Lawler III, *Managerial Attitudes and Performance*, 1968. Copyrighted by Richard D. Irwin, Incorporated.)

TABLE 2

SURVEY SAMPLE BY PREDICTOR VARIABLE

Predictor variable	No. questionnaires distributed	Questionnaires returned	
		No.	%
Base monthly salary			
$1000 and below	245	208	84.8
$1001 to $1200	175	118	67.4
$1201 to $1500	250	166	66.4
$1501 and above	130	78	60.0
Total	800	570	71.3
Job classification			
Programmers and systems analysts	414	305	73.7
All other classifications (Managerial, administrative, engineering, human factors, etc.)	386	270	69.9
Total	800	575	71.9
Performance ratings as judged by supervisor			
High	237	178	75.1
Middle	400	291	72.8
Low	163	105	64.4
Total	800	574	71.8

Hypothesis 4-A.[3] "The higher the perceived probability that pay depends upon job performance factors, the more effort an individual will devote to performing his job effectively [p. 62]."

Hypothesis 4-C. "The relationship between the perceived probability that pay depends upon job performance factors and measures of actual performance and effort will be stronger for those individuals who say their pay is important to them than it will be for those who say their pay is relatively unimportant to them [p. 63]."

Hypothesis 4-D. "The perceived probability that pay depends upon effort will be more highly related to measured actual job performance and effort than will be the perceived probability that pay depends upon quality of job performance [p. 64]."

Hypothesis 4-E. "The more an individual sees his pay as a satisfier, the more effort he will put forth to perform his job effectively [p. 66]."

METHOD

Subjects

All 800 professional employees in a major division of the firm were mailed a survey questionnaire and asked to participate. The final sample consisted of 575 people, or 71.9% of those contacted. Table 2 shows this sample according to three predictor variables used in the survey. It also shows that the sample is representative of the total population from which it was drawn.

Questionnaire

Some of the questions included were from the Porter and Lawler study, and some were of the present authors' own design. An abstract of the questionnaire as applicable to this study appears below.

1. Terminology

   By the terms base salary and base monthly salary, both of which are frequently used throughout the questionnaire, we mean your pay alone. The term compensation, on the other hand, refers to the total income you receive from your position with the company—including base salary and fringe benefits.

2. Predictor Variables

   a. What is your current base monthly salary? $______

   b. What is your current job classification? (Circle the number at the left of your classification).

      1 Assistant Department Manager, Department Manager, Division Vice President

      2 Other Supervisory

      3 Computer Programming

---

[3] Hypotheses numbers are the ones used by Porter and Lawler (1968).

4 Engineering
5 Human Factors
6 Systems Analysis
7 Other

c. Performance rating as judged by supervision.[4]

3. Perceived Self-Ratings

The purpose of this portion of the questionnaire is to determine how you rate yourself relative to others in the company with similar duties. You will be asked to rate yourself for characteristics on a seven-point scale which looks like this:

(low) 1 2 3 4 5 6 7 (High)

Please circle the number on the scale that represents where you feel you stand compared to others. If you think you are low on the characteristic, you would circle the numeral 1. If you think that you are a little less than average as compared with others with similar duties, you would circle the numeral 3, and so on. For each scale, circle only one number.

a. Quality of your work on the job.
(low) 1 2 3 4 5 6 7 (High)
b. Your amount of productivity on the job.
(low) 1 2 3 4 5 6 7 (High)
c. Amount of effort you expend on the job.
(low) 1 2 3 4 5 6 7 (High)

4. Perceived Pay Determiners

In this section there are several characteristics or qualities often used to determine the amount of an individual's base salary. Please indicate how important you feel this company considers these for determining your present base salary. This can be done by using the seven-point scale below each characteristic.

Circle the number on the scale that represents the importance of the characteristic being rated. Low numbers represent less important or unimportant characteristics. If you think this company considers a given characteristic as unimportant in determining the base salary for your position, you would circle numeral 1. If you think it is "just a little" important you would circle numeral 2, and so on.

a. Quality of your work on the job.
(unimportant) 1 2 3 4 5 6 7 (important)
b. Your amount of productivity on the job.
(unimportant) 1 2 3 4 5 6 7 (important)
c. Amount of effort you expend on the job.
(unimportant) 1 2 3 4 5 6 7 (important)

5. In this section are statements designed to measure your attitude about the base monthly salary you are receiving for your present job. Please indicate your agreement or disagreement on the scale below each statement.

a. Perceived Importance of Pay
My base salary level is important to me.

1	2	3	4	5
______	______	______	______	______
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

b. Perception of Pay as a Satisfier
(1) Raises have meant that I was progressing in my work.

1	2	3	4	5
______	______	______	______	______
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

(2) The effort I expend determines the level of my performance.

1	2	3	4	5
______	______	______	______	______
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

---

[4] Employee performance ratings were provided by supervisors prior to the distribution of the questionnaire. The forms were coded in such a way that ratings could be associated with specific individuals before the forms were distributed, but not after they were completed and returned.

(3) Raises I have received were rewards for good performance.

1	2	3	4	5
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

(4) In my job, raises are a form of recognition for a job well done.

1	2	3	4	5
Strongly Agree	Agree	Undecided	Disagree	Strongly Disagree

Attitude Measures

Participants indicated on a 7-point scale (No. 1 on the scale being "unimportant" and No. 7 being "important") the perceived importance of each of nine factors in determining pay. Three factors were used as measures of the importance of job performance. Although not as high as those found by Porter and Lawler, the intercorrelations among these items were high and the questions were thus usable in testing the hypotheses (see Table 3). Some differences may be attributable to the fact that while Porter and Lawler included only managerial personnel in their survey, our study had primarily nonmanagers. Still, the findings suggest considerable homogeneity among factors. Questions used by Porter and Lawler were included to measure the degree to which pay was seen as a satisfier. A 5-point scale was employed, No. 1 being "strongly agree" and No. 5 "strongly disagree." Intercorrelations among the items were high, indicating considerable item homogeneity (see Table 4).

Hypotheses

Since performance factors play a major role in the Porter and Lawler hypotheses regarding the motivational role of pay, hypotheses were tested by computing the predictor variable of supervisory performance rating as a distribution against the remaining variables from the questionnaire. The technique, along with the chi-square significance test, was used to determine both direction and strength of responses. It provided an opportunity to more closely study response patterns than would have been possible through Student's *t* scores as used by Porter and Lawler in testing their hypotheses.

RESULTS

Hypothesis 4-A

Table 5 shows that the perceived probability of pay depending on two of the three job performance factors—work quality and productivity—becomes greater at higher performance levels. A similar trend was obtained for effort, but it was not as strong and was not statistically significant. Even though effort is not seen as leading directly to pay, Table 5 shows that the higher performers see themselves as expending more effort than lower performers. In addition, Table 6 shows that nearly 70% of the population felt that effort determines performance. (The slight difference that occurred on the basis of supervisory performance was not statistically significant.) Hypothesis 4-A is thus supported: As the perceived probability of pay depending on performance becomes greater, the perceived amount of effort expended to perform effectively also becomes greater.

TABLE 3

INTERCORRELATIONS AMONG ITEMS DESIGNED TO MEASURE PERCEPTIONS OF THE IMPORTANCE OF JOB PERFORMANCE FACTORS IN DETERMINING PAY

Item	Productivity			Effort expended		
		Porter/Lawler			Porter/Lawler	
	Survey firm	Government sample	Private industry sample	Survey firm	Government sample	Private industry sample
Quality of work	.70	.80	.71	.47	.66	.50
Productivity				.65	.67	.61
Effort expended						

TABLE 4

INTERCORRELATIONS AMONG ITEMS DESIGNED TO MEASURE THE DEGREE TO WHICH PAY IS SEEN AS A SATISFIER

Item	Item 2			Item 3		
	Survey firm	Porter/Lawler		Survey firm	Porter/Lawler	
		Government sample	Private industry sample		Government sample	Private industry sample
1. Raises have meant that I was progressing in my work.	.62	.64	.61	.59	.49	.52
2. Raises I have received were rewards for good performance.				.64	.61	.65
3. In my job, raises are a form of recognition for a job well done.						

TABLE 5

PERCENTAGE OF RESPONDENTS WHO RATED PERFORMANCE FACTORS AS PAY DETERMINERS VERSUS PERCENTAGE OF RESPONDENTS WHO RATED THEMSELVES ON EFFORT EXPENDED, BY SUPERVISORY PERFORMANCE RATING

Supervisory performance rating	(1 & 2)	(3, 4, 5)	(6 & 7)
	Work quality[f]		
High[a]	3.9	32.6	63.5
Middle[b]	3.8	46.2	50.0
Low[d]	5.8	47.6	46.6
	Productivity[g]		
High[a]	4.5	38.8	56.7
Middle[b]	4.5	52.4	43.1
Low[d]	3.9	54.4	41.7
	Effort[h]		
High[a]	11.8	55.6	32.6
Middle[c]	12.5	58.8	28.7
Low[e]	14.4	57.7	27.9
	Self-ratings of effort expended[i]		
High[a]	0	29.9	70.1
Middle[b]	1.0	48.3	50.7
Low[d]	1.0	51.9	47.1

Note.—Scores in parentheses refer to numbers on the 7-point rating scale used in the questionnaire. Scores of 1 and 2 are classified "unimportant" for the first three performance factors, and "low" for the "self-rating of effort expended." Scores of 6 and 7 are classified "important" for the first three performance factors, and "high" for the "self-rating of effort expended."

[a] $n = 178$.

[b] $n = 290$.

[c] $n = 289$.

[d] $n = 103$.

[e] $n = 104$.

[f] $\chi^2 = 11.39$, $df = 4$, $p = .02$.

[g] $\chi^2 = 10.32$, $df = 4$, $p = .03$.

[h] $\chi^2 = 1.27$, $df = 4$, $p > .85$.

[i] $\chi^2 = 21.71$, $df = 4$, $p < .001$.

Hypothesis 4-C

To test the hypothesis, we divided the sample into those who rated their pay as important and those who rated their pay as unimportant. Within these two groups, we then compared pay and performance-effort perceptions to supervisory performance ratings. Table 7 shows that for those rating pay as important, the higher the supervisory rating the higher the perceived level of performance and effort. For those rating pay as unimportant, the higher the supervisory rating the lower the perceived level of performance and effort. However, the results are not only statistically insignificant for the group rating pay as unimportant, but the large number of expected cell frequencies of less than five indicates that caution should be exercised in interpreting chi-square for this population (see Table 8).

With our data, we were unable to either accept or reject Hypothesis 4-C. Although we

TABLE 6

PERCENTAGE OF RESPONDENTS WHO FEEL THAT EFFORT EXPENDED DETERMINES PERFORMANCE, BY SUPERVISORY PERFORMANCE RATING

Supervisory performance rating[b]	Effort expended determines performance[a]		
	Disagree (4 & 5)	(3)	Agree (1 & 2)
High	16.9	13.5	69.7
Middle	18.9	13.1	68.0
Low	20.0	17.1	62.9

Note.—Scores in parentheses refer to numbers on the 5-point rating scale used in the questionnaire.

[a] $\chi^2 = 1.85$, $df = 4$, $p > .75$.

[b] Ns for the three ratings were 178, 291, and 105, respectively.

TABLE 7

PERCENTAGE OF RESPONDENTS WHO RATED THEIR PERFORMANCE AND EFFORT VERSUS SUPERVISORY PERFORMANCE RATINGS, BY IMPORTANCE OF PAY

Supervisory performance rating[a]	Pay is important[b]			Pay is unimportant[c]		
	Low (1 & 2)	(3, 4, 5)	High (6 & 7)	Low (1 & 2)	(3, 4, 5)	High (6 & 7)
Work quality						
High	0	20.2	79.8	0	50.0	50.0
Middle	0	30.4	69.6	0	43.3	56.7
Low	0	52.6	47.4	0	28.6	71.4
Productivity[d]						
High	0	27.7	72.3	0	75.0	25.0
Middle	0	42.3	57.7	3.4	63.3	33.3
Low	0	55.7	43.3	0	57.1	42.9
Effort[e]						
High	0	28.9	71.1	0	75.0	25.0
Middle	0.8	46.5	52.7	3.4	63.3	33.3
Low	1.0	53.6	45.4	0	28.6	71.4

Note.—Pay is important = those who indicated on the questionnaire that they "agree" or "strongly agree" that pay is important. Pay is unimportant = those who indicated on the questionnaire that they are "undecided," "disagree," or "strongly disagree" that pay is important. Scores in parentheses refer to numbers on the 7-point rating scale used in the questionnaire.

[a] Ns for "Pay is important" = 173, 260, and 97 for high, middle, and low, respectively; for "Pay is unimportant," 4, 30, and 7, respectively.

[b] $\chi^2 = 30.46$, $df = 4$, $p < .001$.

[c] $\chi^2 = .64$, $df = 4$, $p > .95$.

[d] "Pay is important": $\chi^2 = 26.36$, $df = 4$, $p < .001$; "Pay is unimportant": $\chi^2 = .75$, $df = 4$, $p > .90$.

[e] "Pay is important": $\chi^2 = 21.11$, $df = 4$, $p < .001$; "Pay is unimportant": $\chi^2 = 4.08$, $df = 4$, $p > .20$.

found a strong relationship between perceived and rated performance among those who said their pay is important, our data did not enable us to determine a trend or lack of trend for those who rated their pay as unimportant.

Hypothesis 4-D

Table 9 shows that attitudes toward the perceived probability that pay depends on work quality relate more closely to measured actual job behavior than the perceived probability that pay depends on effort. There was more of a tendency for people to see a pay-work quality relationship as performance level became higher. This tendency was much less apparent for the pay-effort relationship, and the small trend that did occur was not statistically significant. We were thus unable to support Hypothesis 4-D with our data.

Hypothesis 4-E

We were able to support the hypothesis with one of three items designed to measure pay as a satisfier ($p = .09$). Table 10 shows that the more pay is seen as a satisfier, the more the amount of effort that is perceived as being expended.

DISCUSSION

The results of our tests of the Porter and Lawler hypotheses regarding the motivational role of pay are summarized below:

Hypothesis 4-A: supported by our research,
Hypothesis 4-C: inconclusive results,
Hypothesis 4-D: not supported by our research,
Hypothesis 4-E: supported by our research.

We found that people will expend more effort the more they see pay as a satisfier and that

TABLE 8

EXPECTED CELL FREQUENCIES OBTAINED IN COMPUTING CHI-SQUARE—SUPERVISORY PERFORMANCE RATING VERSUS PERCEIVED PERFORMANCE LEVEL—RESPONDENTS WHO RATED PAY AS UNIMPORTANT

Supervisory performance rating	Expected cell frequencies		
	Low[a] (1 & 2)	(3, 4, 5)	High (6 & 7)
Work quality			
High	—	1.66	2.34
Middle	—	12.44	17.56
Low	—	2.90	4.10
Productivity			
High	.10	2.54	1.37
Middle	.73	19.02	10.24
Low	.17	4.44	2.39
Effort			
High	.10	2.34	1.56
Middle	.73	17.56	11.70
Low	.17	4.10	2.73

Note.—Scores in parentheses refer to numbers on the 7-point rating scale used in the questionnaire.

[a] The absence of data under this column in "work quality" indicates that no cell frequencies were observed. Therefore, these cells did not contribute to the value of chi-square.

performance (work quality and productivity) is seen as a more important pay determiner than effort. We also found that nearly 70% of our survey participants felt that the amount of effort expended is important in determining their level of performance. It thus appears that although we confirmed two key hypotheses, the function of effort in our survey population is somewhat different than in the theoretical model. In their research, Porter and Lawler found that the more an individual sees his efforts as leading to the attainment of a reward (pay), the more effort he will expend. He first sees his effort as leading to the desired reward, and then proceeds to exert effort which, if role perceptions are accurate, will result in meaningful performance that will in turn lead to the desired reward of pay.

For our survey population, on the other hand, our results led us to conclude that effort is important primarily because it is believed to result in performance. The individual first sees that his performance will lead him to the desired reward. Since he feels that effort expended leads to performance, he will then exert effort which, if role perceptions are accurate, will result in the performance that will attain for him the desired reward of pay.

Porter and Lawler suggested that in a population, such as the one we used, where effort is seen to lead to performance, it may be possible to drop effort from the theoretical model. They said: "Unfortunately we did not directly ask the managers to estimate the probability that effort on their part would result in good performance. If it were found in future studies that this probability was always close to 1.0, then indeed it could be dropped from the model. However, at this point, we still feel that it may

TABLE 9

PERCENTAGE OF RESPONDENTS WHO RATED THE PERCEIVED IMPORTANCE OF WORK QUALITY AND EFFORT AS PAY DETERMINERS, BY SUPERVISORY PERFORMANCE RATING

Supervisory performance rating[a]	Work quality[b]			Effort[c]		
	Unimportant (1 & 2)	(3, 4, 5)	Important (6 & 7)	Unimportant (1 & 2)	(3, 4, 5)	Important (6 & 7)
High	3.9	32.6	63.5	11.8	55.6	32.6
Middle	3.8	46.2	50.0	12.5	58.8	28.7
Low	5.8	47.6	46.6	14.4	57.7	27.9

Note.—Scores in parentheses refer to numbers on the 7-point rating scale used in the questionnaire.

[a] Ns = 178, 290, and 103 under high, middle, and low, respectively, for "work quality" and 178, 289, and 108, respectively for "effort."

[b] $\chi^2 = 11.39$, $df = 4$, $p = .02$.

[c] $\chi^2 = 1.27$, $df = 4$, $p > .85$.

TABLE 10

Percentage of Respondents Who Rated Pay as a Satisfier versus Self-Ratings of Effort Expended

Self-ratings of effort expended[a]	Measures of pay as a satisfier								
	1. Raises have meant that I was progressing in my work.[b]			2. Raises I have received were rewards for good performance.[c]			3. In my job raises are a form of recognition for a job well done.[d]		
	Disagree (4 & 5)	(3)	Agree (1 & 2)	Disagree (4 & 5)	(3)	Agree (1 & 2)	Disagree (4 & 5)	(3)	Agree (1 & 2)
High (6 & 7)	15.7	15.7	68.5	14.9	19.4	65.7	19.2	20.5	60.3
(3, 4, 5)	19.3	15.7	65.1	16.9	22.0	60.2	22.9	26.5	50.6
Low (1 & 2)	29.2	29.2	41.7	25.0	29.2	45.8	33.3	25.0	41.7

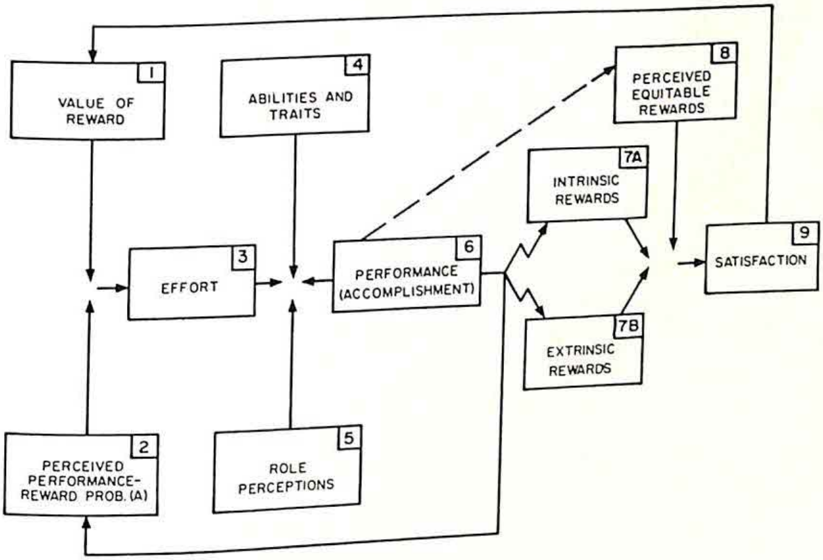
Note.—Scores in parentheses refer to numbers on the 7-point rating scale for "self-rating" and 5-point scale for "pay as a satisfier" used in the questionnaire.

[a] Ns = 464, 83, and 24, respectively, for Item 1; 463, 83, and 24 for Item 2; and 468, 83, and 24 for Item 3.

[b] $\chi^2 = 7.86$, $df = 4$, $p = .09$.

[c] $\chi^2 = 4.58$, $df = 4$, $p > .30$.

[d] $\chi^2 = 6.14$, $df = 4$, $p > .15$.



(A) PERCEIVED EFFORT–REWARD PROBABILITY IN THE ORIGINAL MODEL.

Fig. 2. The Porter and Lawler theoretical model, revised. (Reprinted from L. W. Porter and E. E. Lawler III, *Managerial Attitudes and Performance*, 1968. Copyrighted by Richard D. Irwin, Incorporated.)

be relevant in many situations [p. 88]." We agree with Porter and Lawler that effort is relevant, and our research does not indicate that it should be dropped from the model. Instead, as stated above, its function is different in our population than it is in the theoretical model. The function of effort as it applies to our survey population is shown in Figure 2.

REFERENCE

Porter, L. W., & Lawler, E. E., III. *Managerial attitudes and performance.* Homewood, Ill.: Irwin, 1968.

(Received December 22, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 196–201

CONTINGENCY MODEL OF LEADERSHIP EFFECTIVENESS: SOME EXPERIMENTAL RESULTS[1]

GEORGE GRAEN,[2] JAMES BURDEANE ORRIS, AND KENNETH M. ALVARES

University of Illinois at Urbana

Two parallel experimental studies were performed to replicate and further study the contingency model of leadership effectiveness. The two studies yielded results that were inconsistent with those predicted by the contingency model. These data along with other evidential studies cast doubt on the plausibility of the contingency model of leadership effectiveness.

The contingency model of leadership effectiveness is a theory of leadership effectiveness proposed by Fiedler (1967) in his recent book that summarizes 15 yr. of research on the contribution of leadership style to group performance. This model assumes that the leader's contribution to group performance depends on both the style of leadership (task oriented versus person oriented) and the favorableness of the situation for the leader. The empirical base from which this model was inferred includes a rather broad sampling of actors, behaviors, and contexts, ranging from infantry squads to high school basketball teams (see Fiedler, 1967).

Table 1 presents an eight-cell classification of leadership situations. This classification was produced by dichotomizing on three aspects of the situation: group atmosphere, task structure, and position power. The most important aspect of the situation is assumed to be group atmosphere, which is the leader's evaluation of his members' reactions toward him and his reaction toward the role of leader (Fiedler, 1963). The second aspect, task structure, is the extent the leader is able to control the group's activities based on the extent to which the task is programmable and the number of possible solutions. The least important aspect, position power, is the amount of potential influence given to the leader by virtue of his role occupancy (Fiedler, 1967). The resulting eight octants are hypothesized to represent a dimension of decreasing situational favorableness.

The leadership style dimension varies conceptually from a relationship or person orientation to a task orientation. This dimension of leadership style has been operationalized by an instrument called the "esteem for the Least Preferred Co-worker" (LPC). This instrument asks the respondent to think of the person with whom he can work least well and to rate this person on a number of bipolar adjective scales such as friendly-unfriendly, helpful-frustrating, supportive-hostile, and quarrelsome-harmonious. The lower the leader's esteem for his Least Preferred Co-worker, the more task oriented he is assumed to be, and the greater the LPC score, the greater the relationship orientation.

The dependent variable of the contingency model is the relationship between the leader's style and his group's performance. The usual measure of covariation is the Spearman rank order correlation, rho (McNemar, 1962).

Twelve studies reported by Fiedler (1967) yielded 63 relationships between leadership style and group performance that were used to inductively derive the predictions of the contingency model by a procedure of allowing the median correlation for each octant (see Table 1) to represent the predicted value (Fiedler, 1967, p. 169). The point predictions reveal that the more task-oriented leaders tend to be associated with more productive groups in Octants 1, 2, 3, 8 and with less productive groups in Octants 4 and 5. The opposite is the case for the more person-oriented leaders.

Point predictions from the model are not very interpretable, however, since the variance of correlations within octants is often without an apparent central clustering of values. This variance within octants might be due to the fact that a wide variety of leadership situations were used in developing the model. In order to examine this hypothesis that the variance within octants is due to factors not specified in the model, two closely parallel laboratory experiments were designed which followed as rigorously as possible the prescriptions of the model and the methdology used in its development. In a sense, this is a study of the consistency of the model in that the two studies should yield similar results. Also, these two studies should replicate the contingency model and yield further data points for evidential support or nonsupport of the model.

---

[1] This article has been accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to George Graen, Department of Psychology, University of Illinois, Urbana, Illinois 61801.

In addition to the strong, clearly testable hypothesis of the predicted curve generated by using the results of the 12 antecedent studies, a rather weak and less easily testable hypothesis is also derived by Fiedler from the contingency model. This latter hypothesis merely states that the more task-oriented leader will be associated with more productive groups in very favorable and very unfavorable situations for the leader, and the relationship-oriented leader will be associated with more productive groups in situations that are intermediate in favorability for the leader (Fiedler, 1967, p. 169).

Moreover, the data of the two experiments will be analyzed in a more performance-oriented manner. Instead of looking at the configural patterning of correlation coefficients, group performance will be designated as the dependent variable and the factors of the model (group atmosphere, position power, and leadership style) will be treated as independent variables. In this way, it can be determined not only if the results replicate the contingency model in terms of the pattern of correlations, but, more importantly, if the factors of the model significantly affect performance in terms of mean differences.

Finally, an attempt will be made to clarify the meaning of the LPC score. As mentioned above, a leader with a low LPC score is assumed to behave in a task-oriented manner and a leader with a high LPC score is assumed to act in a person-oriented manner. To examine this hypothesis, a bipolar adjective scale of the same form as the LPC scale was developed to assess the behaviors and attitudes that should be exhibited by high and low LPC leaders. Using this scale, the extent to which the group members perceive the leader as task oriented or person oriented can be determined. In terms of convergent validity, the correlation between the results of this instrument with those of the leader's LPC scale will tend to support or question the LPC score (Cronbach & Meehl, 1955).

TABLE 1

MEDIAN CORRELATIONS FOR THE DEVELOPMENT STUDIES OF THE CONTINGENCY MODEL OF LEADERSHIP

Octant	Situation classification			More effective leadership style (orientation)	*Mdn.* rho	*N*
	Group atmosphere	Task structure	Position power			
1	Good	High	Strong	Task	−.52	8
2	Good	High	Weak	Task	−.58	3
3	Good	Low	Strong	Task	−.33	12
4	Good	Low	Weak	People	.47	10
5	Poor	High	Strong	People	.42	6
6	Poor	High	Weak		—	0
7	Poor	Low	Strong	People	.05	12
8	Poor	Low	Weak	Task	−.43	12

Note.—Median rho (McNemar, 1962) is the median correlation between the Least Preferred Co-worker (LPC) measure of leadership style and group performance. This is the value the model predicts. *N* is the number of relationships included in the calculation of the median correlation.

METHOD

Subjects. The *S*s for both studies were undergraduate students at the University of Illinois who were paid $1.35 for their participation. The first experiment (Exp. I) used 78 *S*s, and the second experiment (Exp. II) used 96 *S*s.

Task. Fiedler operationally defines the degree of task structure by four dimensions developed by Shaw (Fiedler, 1963; Shaw, 1962). These four dimensions are (*a*) goal clarity, (*b*) decision verifiability, (*c*) solution specificity, and (*d*) goal path specificity. Each of these four dimensions is rated on an 8-point scale. Moreover, Fiedler dichotomizes task structure by setting a cutting point of 5.0 on the mean rating over all four scales.

Experiment I used Task No. 23 (high task structure) and Task No. 64 (low task structure) from Shaw's (1962) taxonomy of experimental tasks. The mean ratings derived from the judgments reported in Shaw (1962) for Task No. 23 is 5.86, and that for Task No. 64 is 3.69. Task No. 23 (high structure) is an allocation problem involving the optimum assignment of five operators to five machines, based upon the performance of the operator on each machine. Task No. 64 (low structure) is a human relations problem requiring the selection of the best alternative recommendation from a list of five alternatives.

TABLE 2

COMPARISON OF OBSERVED CORRELATION FROM EXP. I AND EXP. II WITH THOSE PREDICTED FROM THE CONTINGENCY MODEL

Octant	Correlation (rho)		
	Predicted	Observed (Exp. I)	Observed (Exp. II)
1	−.52	.47[a]	−.13
2	−.58	−.41	.18[a]
3	−.33	.46[a]	.02[a]
4	.47	.33	−.08[a]
5	.42	.25	−.52[a]
6	—	−.39	−.43
7	.05	.43	.45
8	−.43	−.33	.44[a]

[a] Sign opposite that predicted.

Experiment II used Task No. 63 (high structure) and Task No. 65 (low structure) from Shaw's (1962) taxonomy. The structure ratings were 5.45 and 3.60 for the two tasks. Task No. 63 (high structure) was the Captain Watts detective problem involving the reconstruction of an incident that resulted in two deaths. Task No. 65 (low structure) was a human relations problem requiring the selection of the best alternative recommendation from a list of five alternatives.

The criterion measure for each task in each experiment was a standardized performance index that combined the quality of the solution and the speed to solution.

Procedure. Both experiments were conducted independently, by different *E*s and at different times, and the results were analyzed independently. However, the procedure for both experiments was identical.

At the beginning of each session, three *S*s were introduced and told that they would be working on problems as a three-man group and that one of them would be designated the leader for each problem. The *S*s were told that it was a study to investigate how groups perform on various types of problems. At this point, each *S* was given the measure of leadership style (LPC) to complete. Next, one member was designated the leader and assigned the authorities and responsibilities of that role. The group was then given the first problem. After completing the first problem, each member of the group completed the group atmosphere scale. This scale consisted of ten 8-point bipolar adjectives of the type used in the LPC. The *S* was asked to describe the atmosphere of the group during problem solving, using these bipolar adjectives. Also, the members filled out the leader behavior questionnaire which consisted of the following six 8-point bipolar adjectives: permissive–strict, requesting–ordering, considerate–rude, participating–managing, passive–active, task-sharing–task-controlling. The left-hand member of each pair refers to high LPC behavior, and the right-hand member refers to low LPC behavior. (The pairs were randomly positioned on the actual questionnaire.) The instrument was scored such that a high score should be associated with high LPC and vice versa.

Then, the second problem was given to the group. After this problem was completed the group atmosphere and leader behavior questionnaires were completed again. Finally, *S*s were debriefed and paid.

Position power of the leader was manipulated according to Fiedler's (1963) checklist: in the strong leader condition, the leader role was assigned superior formal status relative to the members, given special information about the task, and delegated the highest decision-making authority and responsibility. In the weak leader condition, the leader role was one of discussion leader without special information and with decision-making authority and responsibility close to that of the members. Each group worked on both tasks under only one of the two power conditions (strong or weak position power of leader). The task sequence was randomized within power conditions. The selection of leaders for each task was made at random.

RESULTS

The data for each experiment were analyzed by assigning the groups to eight cells of the contingency model by dichotomizing the three situational variables of position power, task structure, and group atmosphere. Position power was defined experimentally as a dichotomy (strong and weak power conditions), and task structure was defined by task selection (structured and unstructured tasks). Finally, group atmosphere was dichotomized by dividing the leaders' total group atmosphere scores at the median (good and poor group atmosphere as perceived by the leader after the experience). This procedure resulted in six groups in each of the first four octants and seven groups in each of the last four octants for Exp. I, and eight groups in each octant for Exp. II. Within each octant, Spearman rank-order correlations (rho) were calculated between leadership style (LPC) and group performance.

The results of these analyses are shown in Table 2. None of the observed correlations for either study reach an acceptable level of statistical reliability. However, the apparent criterion of meaningful relationships for the contingency model merely requires that the observed correlations approximate those predicted at least in terms of direction. For example, in discussing the results of a church-leadership study, Fiedler (1967) states: "While only one of the sixteen (sic) correlations was significant, only one of the eight correlations, namely, .03, was not in the hypothesized direction [p. 119]." This is followed by an interpretation of these correlations that were not significant, but were in the hypothesized direction. Later in discussing the same results, Fiedler concludes:

> The major significance of the church-leadership training study was in the cross-validation and extension of the first two experiments reported in this chapter. The findings obviously held not just for college students but also for mature, experienced leaders; they held for mixed sex groups and over four different creative tasks [p. 120].

Thus the correspondence of the correlations to the hypothesized direction is used as an alternative criterion of meaningful results. The apparent rationale for this is that the need to divide the sample into eight nonoverlapping data sets to produce more homogeneous samples (Fiedler, 1967, pp. 84–85) reduces the sample size to below that necessary to achieve acceptable statistical reliability even with a correlation of moderate strength. This problem of requiring large numbers of experimental units is especially severe when the unit of analysis is a group rather than an individual.

Employing the criterion of the correspondence of the direction of the observed correlations to those predicted (the sign of the median rho's from Table 1), we find in Exp. I that in only two of the seven hypothesized conditions are the observed correlations opposite in sign to those predicted (no prediction was made for Cell 6). Based upon the sign criterion, the advocates of the contingency model might claim support from the results of the first experiment; however, based upon the appropriate statistical criterion, the contingency model hypotheses cannot be supported. The hypothesis for each of the eight samples that the correlation in the population is zero cannot be rejected based upon the observed correlations. Thus, the null hypothesis is accepted for each of the eight samples.

In Exp. II, not only are all of the correlations nonsignificant, but sizable discrepancies appear between observed and predicted correlations for all seven predicted octants. Employing the criterion of the sign of the predicted correlations, only two of the seven correlations are in the hypothesized direction. Thus, the results of Exp. II are damaging both to the hypothesized predicted curve and to the hypothesized inverted-U relationship between the favorability of the situation for the leader and the correlation between leadership style (LPC) and group performance.

To determine the contribution of the factors of the model to group performance, group performance was designated as the dependent variable and position power, group atmosphere, and leadership style (LPC) were treated as independent variables. Moreover, group performance was not assumed to be the same variable over the structured and unstructured tasks. Thus, group performance on each task was considered a different dependent variable. Employing an analysis of variance procedure, leadership style (LPC) was nested within both position power and group atmosphere; both position power and group atmosphere had two levels as defined above, and LPC was dichotomized by cutting at the median within each of the four position power–group atmosphere combinations. This design was used to analyze each of the four task performances. The results of these four analyses of variance show that none of the effects approach an acceptable level of statistical reliability. Following the traditional criterion, the null hypothesis cannot be rejected based upon these test results. The means of group performance are shown in Table 3. Even the differences in means between high and low LPC between the two studies disagree in six of the eight cells.

Finally, the correlation between the leader's LPC score and the mean members' perception of the leader's behavior was $-.01$ for Exp. I and .20 for Exp. II, neither of which was statistically significant.

DISCUSSION

The two studies presented followed as closely as possible the prescriptions of the contingency model and research methodology used in its development. However, neither of the studies supported the model in terms of replicating the predicted curve. Experiment I showed typical contingency model results in terms of the sign of the predicted correlation between leadership style and group performance; however, Exp. II yielded five of seven

TABLE 3

CELL MEANS ON GROUP PERFORMANCE AND CELL FREQUENCIES FOR THE TWO PARALLEL STUDIES

Cell	Octant							
	1	2	3	4	5	6	7	8
$\bar{X}$								
Exp. I								
High LPC	12.04	10.32	11.59	10.19	14.16	8.04	12.03	11.15
Low LPC	10.28	11.54	8.42	8.97	9.74	7.71	9.24	10.62
Difference	1.76	−1.22	3.17	1.22	4.42	.33	2.79	.53
Exp. II								
High LPC	10.80	12.39	10.93	10.65	9.40	9.23	9.92	11.29
Low LPC	11.35	11.58	13.61	9.69	11.46	9.28	10.59	8.62
Difference	−.55	.81	−2.68	.96	−2.06	−.05	−.67	2.67
Frequency								
Exp. I	6	6	6	6	7	7	7	7
Exp. II	8	8	8	8	8	8	8	8

Note.—In six of the octants, the difference between means shows reversed signs for the two parallel studies. In two octants of each study, the sign of the correlation disagrees with that of the difference between means.

correlations with signs opposite to what the contingency model predicts. Perhaps, however, we are too strict in our use of even this most lenient criterion. For example, in describing a study by Hunt (1967), Fiedler (1967) states that "Hunt's results very closely fit the expected curve [p. 176]"; however, Hunt's results include correlations of .60 and −.80 for Octant 3 and correlations of .30 and −.30 for Octant 8. Although Hunt's correlations show both signs, the average (median) for each octant shows the hypothesized sign.

Since there is no acceptable statistical test for comparing the configural patterning of two or more sets of nonsignificant rank-order correlation coefficients, an analysis of variance procedure was employed to determine if the factors of the contingency model had any effect upon group performance. The results of this analysis were completely negative: none of the factors of the model had any influence, either as a main effect or as an interaction. It would be interesting to analyze the other contingency model studies in this manner. It is possible that even in the cases where support is claimed for the contingency model, the effect on performance could still be unreliable, especially when rank-order correlations are used. After all, the dependent variable of interest is group performance, not a rank-order correlation between leader LPC and performance.

It was also found that these two studies did not yield similar results, either in terms of the attempted replication of the contingency model or the mean performance scores within octants. This is a rather damaging result considering the similarity of the experimental conditions, tasks, and *S*s. Consequently, the studies not only lend evidential disproof to the contingency model, but also indicate that it may not be summarizing meaningful and stable relationships.

These studies also failed to contribute any construct validity to the LPC score. A low correlation was found in both studies between the leader's LPC score and the group members' perception of the behaviors that are assumed to define these leadership styles. Perhaps this measure of leadership style (LPC) is the key to the many ambiguous results of studies attempting to replicate the contingency model. A recent review (Graen, Alvares, Orris, & Martella, 1970) summarizes the studies that have been relevant to the evidential validation of the contingency model and points out several conceptual and methodological problems. The two studies presented in the present article, in conjunction with several other studies reviewed by Graen et al. (1970), cast doubt on the plausibility of the contingency model of leadership effectiveness.

REFERENCES

CRONBACH, L. J., & MEEHL, P. R. Construct validity in psychological tests. *Psychological Bulletin,* 1955, **52**, 281–302.

FIEDLER, F. E. A contingency model for the prediction of leadership effectiveness. Urbana, Ill. Group Effectiveness Research Laboratory, University of Illinois, 1963.

FIEDLER, F. E. *A theory of leadership effectiveness.* New York: McGraw-Hill, 1967.

FRENCH, J. R. P., JR., & RAVEN, B. H. Legitimate power, coercive power, and observability in social influence. *Sociometry,* 1958, **32**, 83–97.

GRAEN, G. B., ALVARES, K. M., ORRIS, J. B., & MARTELLA, J. A. Contingency model of leadership effectiveness: Antecedent and evidential results. *Psychological Bulletin,* 1970, **74**, 285–296.

HUNT, J. G. A test of the leadership contingency model in three organizations. Urbana, Ill.: Group Effectiveness Research Laboratory, University of Illinois, 1967. (Also in *Organizational Behavior and Human Performance,* 1967, **2**, 290–308.)

MCNEMAR, Q. *Psychological statistics.* (3rd. ed.) New York: Wiley, 1962.

SHAW, M. E. Scaling group tasks: A method for dimensional analysis. Gainesville, Florida: University of Florida, 1962.

(Received December 5, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 202–204

NOTE ON THE METHODOLOGY OF THE GRAEN, ORRIS, AND ALVARES STUDIES TESTING THE CONTINGENCY MODEL[1]

FRED E. FIEDLER[2]

University of Washington

This note discusses several methodological shortcomings in the Graen, Orris, and Alvares studies of the contingency model. The critique points out that (*a*) the position power manipulation was probably ineffective and at best very weak; (*b*) differences between supposedly structured and unstructured tasks were unnecessarily small, especially since other tasks were readily available; and (*c*) an analysis of variance of leader Least Preferred Co-worker (LPC) scores yielded a highly significant F ratio indicating that some octants in the second study contained primarily high LPC leaders while others contained primarily low LPC leaders. In view of their methodological inadequacies, these experiments do not constitute an adequate test of the contingency model.

Graen, Orris, and Alvares (1971) presented two studies of the contingency model which tested the "hypothesis that the variance within octants is due to factors not specified in the model." This presumably means that the correlation between Least Preferred Co-worker (LPC) and performance is not systematically related to the favorableness dimension. Since the authors do not specify what these other factors might be, the study is simply designed to test a null hypothesis. The authors found no significant relations supporting the contingency model and they concluded that the results constitute "a rather damaging result considering the similarity of the experimental conditions, tasks, and *S*s."

As most of us know all too well, a study can yield nonsignificant results for any number of reasons. Among these, certainly, are insufficiently strong manipulations of important variables as well as inadequate experimental design. A study which claims to disconfirm a theory without at the same time supporting an alternative theory, can, therefore, be expected to meet rather stringent methodological requirements. Graen et al. (1971) recognize this by claiming that their studies "followed as rigidly as possible the prescriptions of the model and the methodology used in its development." The present note will, therefore, confine itself to a discussion of the Graen et al. methodology to point out that these studies have methodological weaknesses which make them inadequate tests of the model.

I should like to leave aside at this time an extended discussion of the need for modifications in the theory. The theory is, in fact, more general than Graen et al. suggest since it postulates primarily a contingent and curvilinear relationship, that is, that the size and direction of the relationship between leader LPC and group performance is moderated by the degree to which the situation gives the leader control and influence over his group. The interested reader is referred to a critique of the contingency model by the writer and his associates (Mitchell, Biglan, Oncken, & Fiedler, 1970), a review of the validation evidence (Fiedler, 1971), as well as a new interpretation of LPC scores (Fiedler, 1970) and a review of selected studies of the contingency model by Graen, Alvares, Orris, and Martella (1970).

The present "note" is primarily a methodological critique of the experimental manipulations and procedures involving the major independent variables in the Graen et al. experiments.

Position power. Graen et al. (1971) experimentally varied position power by giving the randomly chosen leader "superior formal status [and] special information about the task, [and delegating to him] the highest decision making authority and responsibility." This, as ascertained later, was accomplished by talking directly to the leader in the presence of the members, presenting written task instructions to the leader only, "maintaining body orientation and eye contact with the leader," and giving an official timing device to the leader. In the weak position power condition the formal status of the leader was not reinforced, verbal instructions were addressed to the group, and written instructions were given to no one in particular, and the timing device was placed in the center of the table.

---

[1] Published out of turn at the request of the Editor.

[2] Requests for reprints should be sent to the author, Organizational Research, University of Washington, Seattle, Washington 98105.

A number of reasons suggest that this power manipulation is inadequate. First of all, it is difficult to believe that the particular *E* behaviors like looking the leader in the eye and giving the leader of an ad hoc group a timing device and written instructions are sufficiently powerful manipulations to do all the things the authors expected to accomplish. Position power is conceptualized as providing the leader with some real power to give rewards and sanctions. In other words, the leader must have some fate control over his members. It is very difficult to give high position power to a leader in any laboratory situation. Where this was done successfully, it was usually accomplished by using individuals who had some formal position power outside the laboratory. Thus, the Belgian Navy study (Fiedler, 1966) compared petty officers with recruit leaders; a study of ROTC cadets (Meuwese & Fiedler, 1965) and a study of West Point cadets (Skrzypek, 1969) used cadet officers who held higher rank that their members.[3] Moreover, as described in an earlier version of their article, Graen et al. demoted the leader appointed for the first task session to member status in the second session and made another member of the same group the leader in the second session. This procedure, obviously, is likely to dilute the formal leadership power.

Graen et al. have obtained data on the perceived influence of the leader on group performance. These were based on members' ratings of the leader. Graen et al. furnished us average ratings based on the groups within the particular octant. These ratings are based on 8-point scales. They range in Exp. 1 from 6.17 to 5.14 in the high position power condition, and from 4.83 to 3.24 in the low condition; the means for these conditions are 5.57 and 4.32, that is, a difference between means over octants of 1.25.

In Exp. 2, the ranges for the strong position power condition are 6.13–3.50 (sic) for the strong condition and 5.38–4.75 in the weak position. The condition means were 5.29 for the strong, and 5.13 for the weak condition; a difference between means of .16 scale points. It is difficult to believe that the manipulation was successful in the first experiment; it strains credulity to believe that a mean difference of .16 points can be considered evidence showing that the manipulation was successful in the second experiment. Moreover, a *t* test between the eight mean perceived-influence ratings by members for the high condition and the eight means for the low condition yields a *t* of 1.746 which is not statistically significant. When we now consider that each mean per octant is based on a distribution of six to eight groups, it is obvious that the overlap between groups in the high and the low conditions must have been substantial. In light of these considerations, the present author would submit that the position power manipulation in the Graen et al. experiments was clearly inadequate.

Manipulation of task structure. Task structure scores range from a maximum of 8.0 to a minimum of 1.0. The studies reported by Fiedler (1967) average 7.39 for the structured and 3.15 for the unstructured tasks. In contrast, the scores in the Graen et al. studies were 5.86 and 5.45 for the structured tasks and 3.69 and 3.60 for the unstructured tasks. Thus, the structured tasks were less than 1 scale point above the cutting score of 5.0, and the differences between structured and unstructured tasks were only 2.17 and 1.85 scale points. This clearly represents a very weak manipulation as compared with a difference of 4.24 in the original tasks. While it may be argued that the differences between structured and unstructured tasks in some of our studies were no larger, a study which seeks to disconfirm a theory should not rely on marginal experimental manipulations to test a null hypothesis.

Distribution of leader LPC scores. Graen et al. assigned leaders at random within each of the three-man groups, and their LPC scores were determined after constituting the groups. One likely outcome of this procedure is that the leader LPC scores will be unevenly distributed over cells or octants. A sample of groups in which most LPC scores are either extremely high or extremely low is not likely to give results which are predicted by the model since the model assumes a roughly equal distribution of means and variances of LPC scores over the octants (see Fiedler, 1966, 1967).

---

[3] The Skrzypek study, incidentally, yielded correlations almost identical to the predictions of the contingency model, namely, −.43, −.32, .10, .35, .28, .13, .08 and −.33 for Octants 1 to 8, respectively.

Data recently furnished by Graen show that the LPC scores within the various octants had means which ranged from a high of 77.3 to a low of 52.9 in the second study and from 84.7 to 42.7 in the first study. A one-way analysis of variance was performed for each study with the eight octants as cells in the design. The F ratio for the second study was 1.17 indicating that there were no significant differences between means. However, the F ratio for the first study was 3.10 ($p < .01$) showing that the means of leader scores differed markedly from octant to octant. This means that the correlations between leader LPC and group performance are based on biased samples: Some octants contained predominantly high LPC leaders, others contained predominantly low LPC leaders. Since this would reduce the range and variance of LPC scores within octants, it would correspondingly reduce the possibility of obtaining significant correlations.

To sum up, it is rarely possible to insure strong manipulations of all independent variables in a complex experimental design. However, the manipulations of the major independent variables in the Graen et al. studies are inadequate or borderline. Hence, these studies do not constitute critical or very meaningful tests of the contingency model.

REFERENCES

FIEDLER, F. E. The effect of leadership and cultural heterogeneity on group performance: A test of the contingency model. *Journal of Experimental Social Psychology*, 1966, **2**, 237–264.

FIEDLER, F. E. *A theory of leadership effectiveness*. New York: McGraw-Hill, 1967.

FIEDLER, F. E. *Personality, motivational systems, and behavior of high and low LPC persons*. (Organizational Research Tech. Rep. No. 70–12) Seattle, Washington, University of Washington, 1970.

FIEDLER, F. E. Validation and extension of the contingency model of leadership effectiveness: A review of empirical findings. *Psychological Bulletin*, 1971, **76**, in press.

GRAEN, G. B., ALVARES, K. M., ORRIS, J. B., & MARTELLA, J. A. Contingency model of leadership effectiveness: Antecedent and evidential results. *Psychological Bulletin*, 1970, **74**, 285–296.

GRAEN, G. B., ORRIS, J. B., & ALVARES, K. M. Contingency model of leadership effectiveness: Some experimental results. *Journal of Applied Psychology*, 1971, **55**, 196–201.

MEUWESE, W., & FIEDLER, F. E. *Leadership and group creativity under varying conditions of stress*. Urbana, Illinois: Group Effectiveness Research Laboratory, University of Illinois, 1965.

MITCHELL, T. R., BIGLAN, A., ONCKEN, G., & FIEDLER, F. E. The contingency model: Criticism and suggestions. *Academy of Management Journal*, 1970, **13**, 253–267.

SKRZYPEK, G. J. *The relationship of leadership style to task structure, position power, and leader member relations*. (U. S. Military Academy Tech. Rep. No. 34) West Point, New York, 1969.

(Received March 23, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 205-210

CONTINGENCY MODEL OF LEADERSHIP EFFECTIVENESS: SOME METHODOLOGICAL ISSUES

GEORGE GRAEN,[1] JAMES BURDEANE ORRIS, AND KENNETH M. ALVARES

University of Illinois at Urbana

This paper discusses a number of methodological issues concerning empirical studies designed to test several hypotheses derived from the contingency model of leadership effectiveness proposed by Fiedler. These issues are discussed and a number of recommendations are suggested to enhance the information-producing value of future studies designed to test these hypotheses.

Fiedler (1971) raised several selected issues concerning the research methodology of studies designed to test hypotheses associated with his contingency model of leadership effectiveness. We should like to discuss these selected issues as well as several others which emerged during our attempts to understand the empirical phenomena supposedly underpinning this contingency model in the hope that such a discussion will be useful to others performing tests on this model.

Research Strategy and Hypotheses

In contrast to Fiedler's interpretation of the strategy of our two studies, these investigations were designed to test an empirical hypothesis and were not null designs. The alternative explanation for the results of previous contingency studies that was tested by our studies was that the apparent unorderliness of the results of such studies was due to the diversity of behavioral settings investigated. As we stated in our previous article (Graen, Orris, & Alvares, 1971), "the variance of correlations within octants is often without an apparent central clustering of values. This variance within octants might be due to the fact that a wide variety of leadership situations were used in developing the model." The fact that the rejection of the hypothesis, that the results of contingency model studies taken as a set could be explained entirely by the diversity of behavior settings studied, would increase our confidence in the validity of that model, demonstrates that this is not a null hypothesis.

Our approach to testing this hypothesis was to perform two parallel studies within the same behavioral setting to hold the setting effect constant. In this manner, to the extent that the results of the two studies agreed, we could support the reliability of the empirical procedures within this behavioral setting and lend strong support to the validity of the contingency model by decreasing the plausibility of the diversity of settings explanation.

As integral components of our attempt to test the diversity of settings explanation, two versions of the contingency model's central hypothesis were stated and tested in our two studies:

> In addition to the strong, clearly testable hypothesis of the predicted curve generated by using the results of the twelve antecedent studies, a rather weak and less easily testable hypothesis is also derived by Fiedler from the contingency model. This latter hypothesis merely states that the more task oriented leader will be associated with more productive groups in very favorable and very unfavorable situations for the leader, and the relationship oriented leader will be associated with more productive groups in situations that are intermediate in favorability for the leader (Fiedler, 1967, p. 169).

This latter hypothesis represents the curvilinear relationship mentioned by Fiedler as the "more general" hypothesis.

Thus, our studies were designed to test an empirical hypothesis and not simply a null hypothesis. We do not agree, therefore, that studies such as ours be judged according to two different sets of criteria depending upon their results as Fiedler suggests. It does not seem reasonable to apply more "stringent methodological requirements" to a study if its results are damaging to a model than if its results support that model. Although such a procedure may suit the preferences of the theorist,

---

[1] Requests for reprints should be sent to George Graen, Department of Psychology, University of Illinois, Urbana, Illinois 61801.

TABLE 1

CELL MEANS IN PERCEIVED INFLUENCE OF LEADER ON GROUP PERFORMANCE

Group	Octants							
	1	2	3	4	5	6	7	8
Exp. I								
Position power								
Strong	5.83		6.17		5.14		5.14	
Weak		4.33		4.83		3.24		4.86
Exp. II								
Position power								
Strong	5.75		5.63		6.13		3.50	
Weak		5.38		5.00		5.38		4.75

Note.—$p < .04$ that the differences between strong and weak position power means within like quadrants are random.

we feel that it would be dysfunctional for our goal of understanding the empirical phenomenon of interest. For an excellent example of the consequences involved in using this double standard, see the exchange between Rosenthal (1968) and Barber and Silver (1968a, 1968b).

Our attempts to discover some support for the contingency model lead us to even relax the requirement that the observed correlations demonstrate statistical reliability. If the two parallel studies had produced merely similar patterns of nonsignificant correlations, we would have had some basis for hope. However, given the observed dissimilarity of the patterns of correlations between the two studies, one's rationality would be strained severely to find support for the model. A model that claims adequate robustness to have explained leadership effectiveness in a large variety of complex behavioral settings should be expected to produce at least similar patterns of results in two parallel laboratory studies.

Leader's Position Power

Fiedler questioned the treatment effectiveness of our leader's power induction on several grounds. First, he wondered about the kind of special information about the task which could increase the leader's formal power without increasing the task structure. Our procedures to accomplish this objective were to emphasize the expected role behaviors of the designated leader as well as those of the designated members through both the content and the administration of the instructions. For example, the *E* in the high power condition related to the designated leader in both tone and manner as the person in the position of primary responsibility for directing the group's activities. These procedures included: (*a*) addressing all verbal instructions directly to the leader (maintaining appropriate body orientation and eye contact during interactions) in presence of the members, and (*b*) giving the written task instructions and the timing device directly to only the leader. In this manner, the appropriate roles for both the leader and the members were legitimatized in this setting by *E*. In the weak power condition, *E* related to the designated leader as merely another member of the group; verbal instructions were addressed to all three members, and written instructions and timing device simply were placed in the center of the table.

Fiedler also questioned the ability of college students to assume functional roles in problem-solving groups in laboratory situations. If this were the case, much of the small group research, past and present, would be brought into question (Davis, 1969). Moreover, the rewards and sanctions which emerge in the laboratory to produce role acceptance is well documented (Rosenthal, 1966).

In contrast to Fiedler's procedure, in our study we had a somewhat indirect measure of the leader's position power. After each task was completed, the leader filled out a questionnaire which contained a scale asking him how much influence he thought he had over the group's performance. The alternatives were on an 8-point scale ranging from 1, no say, to 8, the final word. The results on this scale are shown in Table 1. As shown in Table 1, seven of the eight differences in means between strong and weak position power within like quadrants are in the expected direction. The one reversal was in Exp. II between Cells 7 and 8. Although this measure probably reflects not only the initial position power of the leader, but also the ability and persuasiveness of the leader relative to that of his members and the leader's perception of the process and outcomes of the group activity, these results, though weak, are generally in the right direction.

Although Fiedler profits from our attempt to evaluate the impact of our leader's power induction through the above indirect measure, his procedure makes no attempt to empirically establish the effectiveness of his position power manipulations from the leader's viewpoint. On the contrary, Fiedler's procedure consists of intuitive judgments by the *E* based upon the expected status and role of the leader. For example in the Meuwese and Fiedler (1967) study, 162 college seniors enrolled in ROTC programs were formed into 54 three-member groups and the member with the highest ROTC rank in a group was designated the leader. The hypothesis that this procedure produced high position power for these leaders was not tested but merely accepted as obvious. As most researchers have learned, intuitively obvious expectations have a way of being disproven by empirical evidence. Based upon this, we submit that our power induction though admittedly weak, probably was representative of some of those employed in previous contingency studies.

Task Structure

Fiedler questioned our selection of experimental tasks as not being different enough. However, in his book (Fiedler, 1967, p. 149) he clearly shows that the cutting point between low and high task structure is 5.0 scale points. Our tasks with scale points of 5.86 and 5.45 for high structure and 3.69 and 3.60 for low structure were above and below this cutting point, respectively. Again, referring to Fiedler's book (pp. 134–141) and considering the antecedent results, 16 correlations as well as two studies by Hawkins (1962) are shown based on studies with task structure scale points between 3.60 and 5.86. Thus, rejecting the above mentioned double standard, our tasks were rather representative of those employed in previous contingency model studies.

Distribution of Leader LPC Scores

Fiedler criticized our procedure of randomly assigning *S*s to conditions on the grounds that this procedure allows for the possibility of finding differences among the conditions on leader's Least Preferred Co-worker (LPC). This criticism apparently is based upon a misunderstanding of the functions of randomization within experimental design. Although this procedure allows for variation of leader LPC scores among the conditions of the design, the expected value of these differences is zero, and the standard deviation of these differences is a function of the standard error of the variable. Thus, the differences that occur among conditions are due to random sampling variation. Using this procedure, the rare event can happen, for example, the occurrence of differences among condition means on leader's LPC that produce an F value with $p < .01$ (Exp. I). However, the more probable events would show differences among the condition means on leader LPC of close to zero, for example, differences among the condition means on leader's LPC of close to zero, that is, differences among condition means on leader's LPC that produce an F value with $p = .33$ (Exp. II). In addition, the difference between the overall means on leader's LPC of the two studies produced an F value with $p = .89$.

Fiedler apparently misinterpreted the implications of this significant F ratio for the differences in leader LPC scores among the eight cells of Exp. I. Fiedler states that this indicates that "the means of leader scores differ markedly from octant to octant." Moreover, he states that this would reduce the variance of LPC scores within octants and thus reduce the possibility of obtaining significant correlations.

The appropriate interpretation of this significant F ratio is that some linear comparison of the eight cell means on LPC shows statistical reliability. It does not indicate that the means differ markedly from octant to octant. This is

TABLE 2

CELL MEANS AND STANDARD DEVIATIONS ON LEADER LPC AND LEADER GROUP ATMOSPHERE

Exp. & octant	N	Leader LPC		Leader GA	
		$\bar{X}$	*SD*	$\bar{X}$	*SD*
Exp. I					
1	6	58.7	15.45	135.2	3.98
2	6	76.7	15.12	129.8	3.85
3	6	66.9	27.53	146.5	6.55
4	6	42.7	19.91	137.8	5.37
5	7	84.7	19.34	106.7	7.50
6	7	48.4	19.92	107.7	14.26
7	7	69.8	12.34	119.4	12.88
8	7	69.7	17.41	123.0	5.81
Exp. II					
1	8	61.8	19.23	146.4	6.74
2	8	52.9	19.63	136.1	6.19
3	8	66.5	29.59	142.3	9.16
4	8	71.8	14.83	135.0	4.33
5	8	56.9	17.29	125.6	13.57
6	8	77.3	16.42	114.6	9.00
7	8	69.5	18.95	105.4	29.24
8	8	62.3	15.77	115.5	12.27

demonstrated by the finding that testing all pairwise differences between means in Exp. I using the Scheffé procedure (Winer, 1962) showed no significant comparisons ($p < .05$). Moreover, as shown in Table 2, the variances on LPC within octants are not noticeably reduced as Fiedler suggests.

Group Atmosphere

In an earlier version of his note, Fiedler objected to our dividing group atmosphere (GA) at the median on the grounds that the differences between good and poor GA probably were too small. Moreover, he suggested that we take groups from the upper and lower third of the GA distribution to assure a larger difference between cells. Fortunately, we have performed an analysis on the results of our two studies, which clearly demonstrates the uselessness of taking successively smaller upper and lower fractions of the GA distribution and in the same process successively reducing the sample sizes within cells. Table 2 shows the correlations within octants for progressively smaller fractions of the GA distribution for each study. Contrary to Fiedler's prediction, the patterns of correlation do not become more supportive of the contingency model. In fact, these results tell us more about the increased variance in the sampling distribution of the rank-order correlation coefficient based upon progressively smaller sample sizes than about the influence of the GA distribution on these correlations. Based upon these results, the present authors suggest that the loss of power (sample size) produced by this procedure is not compensated

TABLE 3

COMPARISON OF THE OBSERVED CORRELATIONS FROM EXP. I AND EXP. II WITH THOSE PREDICTED BY THE CONTINGENCY MODEL USING PROGRESSIVELY LARGER SPLITS ON GROUP ATMOSPHERE

Correlations	GA group		Octant								Σ^b
	Highest	Lowest	1	2	3	4	5	6	7	8	
Predicted			−.52	−.58	−.33	.47	.42	(+)[a]	.05	−.43	
Exp. I	6	6	*.47*	−.41	*.46*	.33	.38	*−.26*	.31	−.23	3
	5	5	*.41*	−.41	*.46*	.36	.30	*−.20*	.20	−.05	3
	4	4	*.63*	−.32	*.40*	*−.32*	.40	*−.40*	.40	−.20	4
Exp. II	8	8	−.13	*.18*	*.02*	*−.08*	*−.52*	*−.43*	.45	*.44*	6
	7	7	−.07	*.18*	−.07	.38	*−.75*	*−.46*	.43	*.15*	4
	6	6	*.49*	*.31*	−.03	.29	*−.71*	*−.37*	.71	−.09	4
	5	5	*.10*	*.60*	*.10*	*−.05*	*−.60*	*−.50*	.90	−.30	6
	4	4	−.11	*.80*	−.80	.95	*−.80*	.00	.80	*.40*	3

Note.—Italicized numbers are of a sign opposite that predicted.

[a] Based upon the contingency model curve this sign should be positive.

[b] Number of signs opposite that predicted per GA group.

by increased orderliness in the pattern of correlations.

Recommendations

We suggest the following procedures be used in studies designed to test the contingency model's predictions regarding the relationship between the leader's LPC and his group's performance within the boundary conditions defined by group atmosphere, task structure, and position power:

1. Position power of the leader should be measured not only by a priori judgment of the proposed procedures, but also by the perception of the leader regarding his influence and/or power on the group process and output. Although this latter procedure normally has not been used (at least not reported) in previous studies, if followed, it would produce data relevant to the hypothesis that the leader's perceived power corresponded to that intended by the investigator. We should take a conservative attitude toward inductions by requiring that treatment checks demonstrate support for our intended inductions.

2. A more acceptable task taxonomy such as the one by Hackman, Jones, and McGrath (1967) be substituted for the one by Shaw (1963). In this way, a taxonomy based upon ratings of the dimensions of group outputs would replace one based upon ratings of the dimensions of task instructions. It is inefficient to perform studies using tasks defined by the latter procedure when tasks defined by the former are available. Moreover, Hackman is now extending his excellent taxonomy to nonlaboratory tasks.

3. Leadership style should be monitored as well as indexed by LPC and various other measures. The lack of agreement between the leadership style perceived by group members and that indexed by LPC in the studies by Graen and his associates demonstrates the need for monitoring the leader's behavior. Before we attribute variance in group performance to the leader's style, we should show data that the leader's behavior was as assumed. This monitoring of the leader's behavior should help to dispel the mystery surrounding the concept of the LPC.

4. Group atmosphere should be manipulated and monitored as part of the design and not merely measured after the fact. Measuring GA after the fact may confound it with other group outcomes such as the leader's perceptions of the group task processes and outcomes (group atmosphere is as perceived by the leader). Thus, the manipulation of GA would allow unconfounded interpretation of the contributions of this factor.

5. The dependent variable of a rank-order correlation between LPC and group performance should not be accepted as meaningful unless an appropriate statistical test shows it to be statistically reliable at an acceptable level. The consequences in the case of the contingency model research of not employing this traditional convention is shown in the review of these studies (Graen, Alvares, Orris, & Martella, 1970).

6. In addition to the usual analysis of correlations within octants, an appropriate analysis of variance design should be used to test the significance and show the strength of association of each source of variation on raw group performance. Moreover, a priori contracts on the means produced by this analysis should be tested (Winer, 1962).

REFERENCES

BARBER, T. X. & SILVER, M. J. Fact, fiction, and the experimenter bias effect. *Psychological Bulletin*, 1968, **70**, (6, Pt. 2). (a)

BARBER, T. X., & SILVER, M. J. Pitfalls in data analysis and interpretation: A reply to Rosenthal. *Psychological Bulletin*, 1968, **70** (6, Pt. 2). (b)

DAVIS, J. H. *Group performance.* Reading, Massachusetts: Addison-Wesley, 1969.

FIEDLER, F. E. *A theory of leadership effectiveness.* New York: McGraw-Hill, 1967.

FIEDLER, F. E. A note on the methodology of the Graen, Orris, and Alvares studies testing the contingency model. *Journal of Applied Psychology*, 1971, **55**, 202–204.

GRAEN, G., ALVARES, K. M., ORRIS, J. B., & MARTELLA, J. A. Contingency model of leadership effectiveness: Antecedent and evidential results. *Psychological Bulletin*, 1970, **74**, 285–296.

GRAEN, G., ORRIS, J. B., & ALVARES, K. M. Contingency model of leadership effectiveness: Some experimental results. *Journal of Applied Psychology*, 1971, **55**, 196–201.

HACKMAN, J. R., JONES, L. E., & MCGRATH, J. E. A set of dimensions for describing the general properties of group-generated written passages. *Psychological Bulletin*, 1967, **70**, 379–390.

HAWKINS, C. A. Studies of factors mediating a relationship between leader role behavior and group productivity. Unpublished doctoral dissertation. University of Minnesota, 1962.

MEUWESE, W., & FIEDLER, F. E. Leadership and group creativity under varying conditions of stress. In F. E. Fiedler (Ed.), *A theory of leadership effectiveness.* New York: McGraw-Hill, 1967. (Also published by Group Effectiveness Laboratory, 1965).

ROSENTHAL, R. *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts, 1966.

ROSENTHAL, R. Experimenter expectancy and the reassuring nature of the null hypothesis decision procedure. *Psychological Bulletin*, 1968, **70** (6, Pt. 2).

SHAW, M. E. *Scaling group tasks: A method for dimensional analysis.* Gainesville, Florida: University of Florida, 1963. (Mimeo).

WINER, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received May 20, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 211–217

REGRESSION ANALYSIS OF THREE VARIATIONS OF THE TWO–FACTOR THEORY OF JOB SATISFACTION[1]

CHARLES L. HULIN[2]
University of Illinois at Urbana

AND

L. K. WATERS
Ohio University

Three distinct variations of the two-factor theory were tested by reanalyzing data previously published by Waters and Waters. No support for any of these three frequently assumed versions was found. The results of this and several prior studies were explained by assuming the existence of two main effects: The first is that intrinsic job factors are generally more potent for predicting overall job attitudes, and the second is that overall job satisfaction is more predictable than overall job dissatisfaction. Neither effect requires the assumptions of the discontinuity of satisfaction and dissatisfaction or the assumptions of dimensional differences between intrinsic and extrinsic job factors.

The two-factor theory (Herzberg, 1966; Herzberg, Mausner & Snyderman, 1959) has generated an impressive number of studies since its appearance over 10 yr. ago. In spite of the number of studies which have appeared in these 10 yr., King (1970) has pointed out that no fewer than five different versions of the two-factor theory have either been implicitly assumed or explicitly stated by researchers in designing their studies. Further, King has argued, two of the five versions of the two-factor theory have not been adequately tested since apparent support may reflect little more than method variance. The existence of these five versions represents a serious problem since data which may support one variation of the theory may not support a different variation.

Three of the versions of the two-factor theory identified and discussed by King (1970) are under consideration in this article. These three versions are summarized briefly below.

Theory 1. This version of the theory states that variation in all intrinsic variables combined contributes more to variation in job satisfaction than it does to variation in job dissatisfaction, and variation in all extrinsic variables combined contributes more to variation in job dissatisfaction than to variation in job satisfaction. This is the version assumed by Whitsett and Winslow (1967) for their review of the Dunnette, Campbell, and Hakel (1967) study.

Theory 2. This theory predicts that variation in all intrinsic variables combined contributes more to variation in job satisfaction than does variation in all extrinsic variables combined, and variation in all extrinsic variables combined contributes more to variation in job dissatisfaction than does variation in all intrinsic variables combined. Herzberg (1966) implied that Theory 2 is the basic theory since he states:

> Once again, ... the divergent trends for motivators and hygiene factors serve to verify the basic theory as these factors are involved in positive and negative job attitude sequences [p. 101].

Theory 3. This theory states that each intrinsic variable contributes more to job satisfaction than it does to job dissatisfaction, and each extrinsic variable contributes more to job dissatisfaction than it does to job satisfaction. Theory 3 is a stronger version of Theory 1. Support for Theory 3 implies support for Theory 1. On one occasion, Herzberg (1966) implied that Theory 3 was the basic two-factor theory since he stated:

> The chart shows that of the 51 significant differences reported for the six motivator factors, every one was in the predicted direction. For the 57 significant hygiene factors, 54 were in the predicted direction. In sum, then, the predictions from the theory were wrong in less than 3 percent of the cases [p. 125].

These three versions of the two-factor theory may be summarized as follows: Let a_i, b_i, c_i, and d_i represent either correlations between job factors and overall attitudes or frequencies of reporting critical incidents associated with either positive or negative job satisfaction. That is, if a_i represents the correlation between job factor i and overall satisfaction, then b_i represents the correlation between the same factor and overall dissatisfaction. Likewise, if a_i represents the frequency of reporting factor i being associated with overall satisfaction, then b_i represents the frequency of reporting that the same factor was associated with overall dissatisfaction.

---

[1] The authors would like to thank Thomas Taber and Carrie W. Waters for their assistance and comments on this article.

[2] Requests for reprints should be sent to Charles L. Hulin, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

	Overall satisfaction (y_1)	Overall dissatisfaction (y_2)
Intrinsic factors	a_1 a_2 $\vdots$ a_n	b_1 b_2 $\vdots$ b_n
Extrinsic factors	c_1 c_2 $\vdots$ c_n	d_1 d_2 $\vdots$ d_n

Theory 1 states that $\sum_{i=1}^{n} a_i > \sum_{i=1}^{n} b_i$ and $\sum_{i=1}^{n} d_i > \sum_{i=1}^{n} c_i$. In the case of a correlational study, Theory 1 would predict that the multiple correlation based on the intrinsic factors, a_i, would be greater than the corresponding multiple correlation based on b_i ($R_{a_1a_2\cdots a_n,y_1} > R_{b_1b_2\cdots b_n,y_2}$). Theory 1 would also predict that $R_{d_1d_2\cdots d_n,y_2} > R_{c_1c_2\cdots c_n,y_1}$.

Theory 2 predicts that $\sum_{i=1}^{n} a_i > \sum_{i=1}^{n} c_i$ in the case of a critical incident study, and $R_{a_1a_2\cdots a_n,y_1} > R_{c_1c_2\cdots c_n,y_1}$ in the case of a correlation study. Theory 2 also predicts that $\sum_{i=1}^{n} d_i > \sum_{i=1}^{n} b_i$, in the case of a critical incident study and $R_{d_1d_2\cdots d_n,y_2} > R_{b_1b_2\cdots b_n,y_2}$ in the case of a correlational study.[3]

Theory 3 predicts that $a_i > b_i$ for all a_i, and b_i and $d_i > c_i$ for all d_i and c_i. Note again that Theory 3 is a stronger version of Theory 1 and support for Theory 3 implies support for Theory 1.

King (1970) concludes that Theory 1 has been supported by critical incident studies and subject-coded studies but has not been tested adequately in studies where the determinants of satisfaction and dissatisfaction were obtained by means other than self-report. A number of studies apparently testing deductions of Theory 1 were discarded by King and earlier by Hulin and Smith (1967) because these discarded studies measured overall satisfaction and dissatisfaction on a single scale rather than following the assumptions of the two-factor theory and assuming satisfaction and dissatisfaction were on two separate and distinct continua. King, and Hulin and Smith contend that if a theory denies the existence of a single dimension of overall satisfaction/dissatisfaction and makes no predictions about it, it is not appropriate to test the theory using a single satisfaction/dissatisfaction continuum.

Theory 2 has not been tested adequately in studies other than critical incident studies according to King (1970).

Theory 3 is supported only by critical incident studies. Support for this theory appears to reflect method variance. King (1970) argues that only one study (Hulin & Smith, 1967) adequately tests this theory using a method different from critical incident studies. This study reports data which fail to support the predictions of Theory 3. Recently, Waters and Waters (1969) reported data which also failed to support Theory 3. Unfortunately, Hulin and Smith did not gather data appropriate to test Theories 1 and 2 (according to King), and Waters and Waters analyzed their data to test Theory 3 but not Theories 1 and 2.

This article reports a reanalysis and reinterpretation of the Waters and Waters (1969) data to test Theories 1 and 2. Specifically, two methods of measuring satisfaction (the Job Description Index [JDI] scales and a set of Likert scales) were used to assess satisfaction with various job factors. Part and multiple correlations between these factors and measures of overall satisfaction and overall dissatisfaction were obtained in order to evaluate the linear association between the job factors and overall job attitudes.

METHOD

The respondents in this study were 160 nonsupervisory female employees in one regional office of a national insurance company. The employees ranged in age from the late teens to early 60s and all were at least high school graduates. Immediate supervisors of all employees sampled were also females.

---

[3] This is a somewhat simplified statement of an appropriate test of Theory 2. See the discussion in the Results section for some caveats.

A job attitude questionnaire was administered to small groups of employees by the second author during a single working day. Respondents were assured that their individual responses would not be made known to the company. Names were requested but employees were given the option of not responding to that item if they "felt uncomfortable" doing so. Approximately 17% did not give their names. Other information (job title, department, etc.) obtained on the questionnaire probably would have been sufficient to identify the respondent.

The job attitude scales were presented in booklet form and consisted of separate overall satisfaction and dissatisfaction scales (always the first two scales, order randomized), an overall satisfaction/dissatisfaction scale, the five scales of the JDI, and a list of 11 job factors (arranged in alphabetical order) to be rated on a satisfaction/dissatisfaction scale. Ratings of satisfaction/dissatisfaction (both overall and for specific job factors) were made on a 12-point anchored scale and the separate satisfaction and dissatisfaction ratings on 7-point scales which consisted of the appropriate 6 points of the 12-point satisfaction/dissatisfaction scale plus a seventh alternative (not satisfied or not dissatisfied). Other data were gathered, but only the above data are analyzed in this article. See Waters and Waters (1969) for details of these other data.

RESULTS

Tests of Theory 3

The explication of the five versions of the theory make the distinctions between them clear. It is not clear, however, how one would test the deductions of the theories. Waters and Waters (1969) and Hulin and Smith (1967), for example, attempted to test the deductions of Theory 3 by comparing the correlations between measures of satisfaction with various job factors and overall satisfaction or overall dissatisfaction. Such a procedure ignores confounding caused by the covariance between the measures of satisfaction with the different job factors. Partial correlations between measures of satisfaction with the job factors and overall satisfaction, or standard partial regression weights relating satisfaction with job factors to overall satisfaction might be appropriate methods which would take account of the covariance terms. Theory 3 was tested in this article by means of the regression of each of the overall measures on the set of independent variables. The standard partial regression weights indicating the relationship between each of the 11 job variables measured by Likert scales on the one hand and overall job satisfaction and overall job dissatisfaction on the other are presented in Table 1. The 11 job variables are grouped so that the intrinsic variables appear first, followed by the extrinsic variables, and finally by the variable "salary" which is unclassified in this table. According to Theory 3 each intrinsic variable should contribute more to overall satisfaction than to overall dissatisfaction. Each extrinsic variable should contribute more to overall dissatisfaction than to overall satisfaction. Thus, according to Theory 3, the first five weights in Column 1 of Table 1 should each be larger than the corresponding entries in Column 2 of Table 1. Only the weights assigned to recognition and achievement are in line with the prediction of Theory 3.[4]

TABLE 1

STANDARD PARTIAL REGRESSION WEIGHTS RELATING INTRINSIC AND EXTRINSIC JOB VARIABLES TO OVERALL JOB SATISFACTION AND OVERALL JOB DISSATISFACTION: LIKERT SCALES

Variables	Overall satisfaction	Overall dissatisfaction
Intrinsic variables		
Responsibility	.08	.08
Work itself	.17	.19
Growth and advancement	−.11	.04
Recognition	−.03	−.13
Achievement	.26	.13
Extrinsic variables		
Competent supervision	.18	.23
Considerate supervision	.07	.07
Company policies and practices	.18	.09
Co-workers	−.03	−.10
Working conditions	.06	.14
Unclassified variables		
Salary	.16	.02
R^2	.43	.30

Note.—$N = 160$.

---

[4] It seems unlikely that the lack of predictability could be due to an inadequate sampling of job factors to use as predictor variables. The five JDI scales were chosen so as to be consistent with the important factors of job satisfaction which have been found in previous studies. The 11 job variables measured by the Likert scales in this study appear to do a more than adequate job of sampling.

TABLE 2

STANDARD PARTIAL REGRESSION WEIGHTS RELATING INTRINSIC AND EXTRINSIC JOB VARIABLES TO OVERALL JOB SATISFACTION AND OVERALL JOB DISSATISFACTION: JOB DESCRIPTION INDEX

Variables	Overall satisfaction	Overall dissatisfaction
Intrinsic variables		
Work	.52	.36
Promotions	−.02	.03
Extrinsic variables		
Supervision	.15	.19
Co-workers	.05	−.07
Unclassified variables		
Pay	.15	.12
R^2	.44	.25

Note.—N = 160.

Also, according to Theory 3 the weights for each of the extrinsic variables should be larger in Column 2 than they are in Column 1. Again, two of the five, competent supervision and working conditions, meet this criterion.

A parallel analysis of the relationship between the job variables measured by the JDI and overall satisfaction and dissatisfaction is presented in Table 2. Theory 3 would predict that the first two weights in Column 1 should be larger than the corresponding weights in Column 2. One, work satisfaction, meets this criteria. Each of the weights for the extrinsic variables should be larger in Column 2 than it is in Column 1. One of the two, supervision, meets this criterion.

No tests of significance of the differences between the regression weights were computed, since only 40% of the predictions of Theory 3 were correct in Table 1, and only 50% of the predictions were correct in Table 2. A theory should do better than chance before we compute significance levels of differences. These data support the results and conclusions of Hulin and Smith (1967) and Waters and Waters (1969) which were based on simple correlational analyses.

Tests of Theory 2

Theory 2 is somewhat more difficult to test directly than Theory 3. Ideally, we would like to obtain the regression of each dependent variable on the 11 job variables (as was done in Tables 1 and 2) and convert the standard partial regression weights associated with each set of independent variables (intrinsic and extrinsic) into estimates of the percentage of variance of the dependent variable accounted for by each set. Unfortunately, this can only be done in the case where the covariances between the independent variables are zero. Such is not the case with these data. Two alternative strategies were followed. First, the regressions of overall satisfaction and dissatisfaction on the intrinsic variables and on the extrinsic variables were computed. The squared multiple correlations which resulted from these four analyses, using the Likert scales as the independent variable, are given in Table 3. The results of the same analyses using the JDI scales as the independent variables are given in Table 4. It should be noted that the Likert scale measuring satisfaction with salary and the JDI scale measuring pay satisfaction were considered to be extrinsic variables in all subsequent analyses. This is in keeping with Herzberg's original classification. Separate analyses were done dropping both pay (salary) scales entirely. The results were not changed appreciably.

According to Theory 2 the intrinsic variables should account for a greater proportion of the variance of overall satisfaction than the extrinsic variables do. The results given in Tables 3 and 4 support this hypothesis. Theory 2 also predicts that the extrinsic variables should account for a greater proportion of the variance of overall dissatisfaction than do the intrinsic variables. The results given in Tables 3 and 4 do not support this prediction. Thus,

TABLE 3

PROPORTIONS OF VARIANCE (R^2) OF OVERALL SATISFACTION AND OVERALL DISSATISFACTION ACCOUNTED FOR BY INTRINSIC AND EXTRINSIC VARIABLES CONSIDERED SEPARATELY: LIKERT SCALES

Variable	Overall satisfaction	Overall dissatisfaction
Intrinsic	.34	.23
Extrinsic	.29	.22

Note.—N = 160.

TABLE 4

PROPORTIONS OF VARIANCE (R^2) OF OVERALL SATISFACTION AND OVERALL DISSATISFACTION ACCOUNTED FOR BY INTRINSIC AND EXTRINSIC VARIABLES CONSIDERED SEPARATELY: JOB DESCRIPTION INDEX SCALES

Variable	Overall satisfaction	Overall dissatisfaction
Intrinsic	.39	.21
Extrinsic	.35	.15

Note.—N = 160.

TABLE 5

PROPORTIONS OF VARIANCE OF OVERALL SATISFACTION AND DISSATISFACTION UNIQUELY CONTROLLED BY INTRINSIC AND EXTRINSIC VARIABLES: LIKERT SCALES

Variable	Overall satisfaction	Overall dissatisfaction
Intrinsic	.14	.08
Extrinsic	.09	.07

Note.—N = 160.

two of the four predictions made by Theory 2 are supported in Tables 3 and 4.

It is not entirely clear, however, that the analyses summarized in Tables 3 and 4 are appropriate. In these analyses, the intrinsic and extrinsic variables were not allowed to compete for the same pool of variance since the regressions of overall job attitudes on the intrinsic and extrinsic job factors were obtained separately. It is clear that there is considerable overlap in the variance accounted for by the two sets of independent variables. For example, the squared multiple correlation of all of the variables predicting overall satisfaction is .43, yet the sum of the squared multiple correlations of the two sets of variables predicting overall satisfaction is .63. An alternative analysis would be to obtain the proportion of variance of the overall satisfaction and dissatisfaction measures which are uniquely (given the independent variables at our disposal) controlled by each set of independent variables. In order to obtain these values, the multiple correlations between all predictor variables combined and each overall measure were obtained. The multiple correlations between each subset and predictor variables (intrinsic and extrinsic) taken separately were obtained next. These values are sufficient to compute the amount of variance of the dependent variable uniquely accounted for by the intrinsic or extrinsic variables. For example, the difference between R^2 using all variables to predict overall satisfaction and R^2 using only the extrinsic variables is the proportion of variance of overall satisfaction uniquely accounted for by the intrinsic variables. These proportions of variance are given in Table 5 (for the Likert scales) and Table 6 (for the JDI scales).

Theory 2 would predict that the proportion of variance of overall satisfaction uniquely accounted for by the intrinsic variables would be greater than the variance of overall satisfaction accounted for by the extrinsic variables. This prediction is supported both in Table 5 and Table 6. Theory 2 would also predict that the proportion of variance of overall dissatisfaction uniquely accounted for by extrinsic variables would be greater than that accounted for by intrinsic variables. This prediction is supported in neither Table 5 nor Table 6. Two of the four predictions made by Theory 2 are supported in Tables 5 and 6. Thus in Tables 3, 4, 5, and 6 only one-half of the predictions made by Theory 2 are supported. A random model would do as well.

Tests of Theory 1

Theory 1 predicts that intrinsic variables account for a greater proportion of the variance of overall satisfaction than they do of overall dissatisfaction. Likewise, extrinsic variables should account for a greater proportion of the variance of overall dissatisfaction than

TABLE 6

PROPORTIONS OF VARIANCE OF OVERALL SATISFACTION AND DISSATISFACTION UNIQUELY CONTROLLED BY INTRINSIC AND EXTRINSIC VARIABLES: JOB DESCRIPTION INDEX SCALES

Variable	Overall satisfaction	Overall dissatisfaction
Intrinsic	.09	.10
Extrinsic	.05	.04

Note.—N = 160.

they do of overall satisfaction. Tables 3, 4, 5, and 6 present the data necessary to test Theory 1. Three of the eight predictions made by Theory 1 are supported. Four of the eight would be chance. It did not seem necessary to compute significance levels.

DISCUSSION

The data reported in this article would seem to meet all of the criteria discussed by King (1970) for testing three of the two-factor theories. That is, overall satisfaction and dissatisfaction were assessed on separate scales; a wide range of individual job factors was considered; the analyses were adequate to test the different versions of the theory. Fewer than 50% of the predictions of the three versions of the two-factor theory were supported. No version of the theory made more than 50% correct predictions. We again arrive at the oft-stated conclusion: The two-factor theories receive no support when tested using data generated by methods other than the critical incident or *S*-coded techniques.

Two trends in the data analyzed in this article are noteworthy, aside from the specific tests of the theories. All of the data in Tables 3–6, with the exception of one reversal (i.e., Row 1, Table 6), can be accounted for by assuming the existence of two main effects. The first main effect is that intrinsic variables are generally more potent and account for more of the variance in workers' overall attitudinal responses to their jobs than extrinsic variables. Independent support for this hypothesis may be found in Hulin and Smith (1967), Waters and Waters (1969), and Graen (1966). The second main effect is that overall satisfaction is more predictable than overall dissatisfaction. Whether this is due to reliability, validity, or choice of independent variables is unknown. Independent support for this hypothesis may be found in Hulin and Smith (1967), Graen and Hulin (1968), and Graen (1966). These two main effects would result in a large proportion of the variance of overall satisfaction being accounted for by intrinsic characteristics, a small proportion of the variance of overall dissatisfaction accounted for by extrinsic variables, with the other two cells being intermediate in size (see Tables 3–6). We need assume nothing about the continuity or discontinuity of overall responses to the job. Neither do we need assume anything about the qualitative differences between intrinsic and extrinsic variables. Two simple main effects, one due to the relative potency of the intrinsic and extrinsic variables and one due to psychometric characteristics are sufficient to account for these data. It appears that a two-factor theory may account for much of the data generated by Herzberg's theory. It does not appear to be the two-factor theory which was originally proposed.

It should be noted that only three versions of the two-factor theory were investigated in this study. However, King (1970) has concluded that Theory 4 and Theory 5, being stronger versions of the first three versions, have received no support in the literature. Thus, regardless of which particular version of the two-factor theories one assumed was "the theory," little support can be found for it in the world of data.

Finally, we should end on a note of caution. The problem of multicollinearity complicates the interpretation of Tables 1 and 2. Gordon (1968) has demonstrated the theoretical and logical absurdity of using multiple measures of the same construct in a regression analysis. Similar problems of interpretation are caused when a number of linearly related variables are included in the same regression analysis. The negative β weights associated with growth and advancement, recognition, and co-workers, even when these variables have positive correlations with the criteria, indicates the multicollinearity which exists among the 11 Likert scales and further indicates that these 11 scales are redundant for predicting overall attitudinal responses to the job. Therefore, Tables 1 and 2 should be approached with all due respect for the problem of unstable βs. If only one intrinsic variable and one extrinsic variable were included in the analyses, this problem would be removed. Unfortunately, this procedure would not adequately test the theory.

REFERENCES

DUNNETTE, M. D., CAMPBELL, J. P., & HAKEL, M. D. Factors contributing to job satisfaction in six occupational groups. *Organizational Behavior and Human Performance*, 1967, **2**, 143–174.

GORDON, R. A. Issues in multiple regression. *The American Journal of Sociology*, 1968, **73**, 592–616.

GRAEN, G. B. Addendum to "An empirical test of the Herzberg two-factor theory." *Journal of Applied Psychology*, 1966, **50**, 551–555.

GRAEN, G. B., & HULIN, C. L. Addendum to "An empirical investigation of two implications of the two-factor theory of job satisfaction." *Journal of Applied Psychology*, 1968, **52**, 341–342.

HERZBERG, F. *Work and the nature of man.* Cleveland: World, 1966.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. *The motivation to work.* (2nd ed.) New York: Wiley, 1959.

HULIN, C. L., & SMITH, P. C. An empirical investigation of two implications of the two-factor theory of job satisfaction. *Journal of Applied Psychology*, 1967, **51**, 396–402.

KING, N. A clarification and evaluation of the two-factor theory of job satisfaction. *Psychological Bulletin*, 1970, **74**, 18–31.

WATERS, L. K., & WATERS, C. W. Correlates of job satisfaction and job dissatisfaction among female clerical workers. *Journal of Applied Psychology*, 1969, **53**, 388–391.

WHITSETT, D. A., & WINSLOW, E. K. An analysis of studies critical of the motivator-hygiene theory. *Personnel Psychology*, 1967, **20**, 391–415.

(Received February 16, 1970)

Manuscripts Accepted for Publication in the *Journal of Applied Psychology*

Anxiety as a Function of Intrinsic-Extrinsic Job Orientation: The Presence or Absence of Observers, and Task Difficulty: Shoukry D. Saleh*: Department of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada.

Effects of Contingent and Noncontingent Reward on the Relationship between Satisfaction and Task Performance: David J. Cherrington, H. Joseph Reitz,* and William E. Scott, Jr.: Business 650, Braduate School of Business, Indiana University, Bloomington, Indiana 47401.

Spatial and Temporal Information as Cues in a Time-Sharing Task: Peter W. Dorfman, and Irwin L. Goldstein*: Department of Psychology, University of Maryland, College Park, Maryland 20742.

Preinvention Effects of Mediation Versus Arbitration: Douglas F. Johnson,* and Dean G. Pruitt: Management Research Center, College of Business Administration, University of Rochester, Rochester, New York 14627.

Negro-White Differences in Validity of Employment and Training Selection Procedures: Summary of Research Evidence: Virginia R. Boehm*: Division of Employment, New York State Department of Labor, 370 Seventh Avenue, New York, New York 10001.

Work Sampling for Personnel Selection: James E. Campion*: Department of Psychology, University of Houston, Cullen Boulevard, Houston, Texas 77004.

Defense Mechanisms as an Alternate Explanation of Herzberg's Motivator-Hygiene Results: H. Randolph Bobbitt, Jr.,* and Orlando Behling: Faculty of Management Sciences, Ohio State University, 1775 South College Road, Columbus, Ohio 43210.

Size, Style, and Vertical Spacing in the Legibility of Small Typefaces: E. C. Poulton, Medical Research Council, MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, England.

Predictive Validity of Vocational Interests within Nonprofessional Occupations: Alan W. Lau,* and Norman M. Abrahams: Naval Personnel and Training Research Laboratory, Department of the Navy (Code 52) San Diego, California 92152.

Effectiveness of Pay Incentives under Variable Ratio and Continuous Reinforcement Schedules: Gary Yukl,* Kenneth N. Wexley, and James D. Seymore: Department of Psychology, University of Akron, 302 East Buchtel Avenue, Akron, Ohio 44304.

Racial Differences in Job Attitudes: John W. Slocum, Jr.,* and Robert H. Strawser: College of Business Administration, Pennsylvania State University, 120 Boucke Building, University Park, Pennsylvania 16802.

Influence of Community and Plant Characteristics on Job Attitudes of Manual Workers: Ray Wild,* and T. Kempner: Production Programme Management Centre, University of Bradford, Emm Lane, Bradford 9, England.

Success and Preference: A Replication and Extension: Samuel H. Osipow*: Department of Psychology, College of Social and Behavioral Sciences, Ohio State University, 1945 North High Street, Columbus, Ohio 43210.

Subordinate's Assigned Level of Discretion as a Function of Leader's Personality and Situational Variables: Ahmed Sakr Ashour,* and George England: Department of Management and Marketing, School of Business, Fresno State College, Fresno, California 93710.

Factors in the Job Involvement of Middle Managers: Winston R. Schwyhart,* and Patricia Cain Smith: Naval Personnel and Training Research Laboratory, Personnel Measurement Department, San Diego, California 92152.

---

* Asterisk indicates author for whom address is supplied.

Journal of Applied Psychology
1971, Vol. 55, No. 3, 218–222

EXPECTANCIES AS DETERMINANTS OF PERFORMANCE

ABRAHAM K. KORMAN[2]

New York University

This research reports five studies which all support the general proposition that high expectancies of competence by others are positively related to performance. Two of the studies involved laboratory research where the expectancies emanated from peers while the other three involved field research in different types of working situations where the expectancies were those of organizational leadership vis-à-vis subordinate performance. Intelligence and environmental variables were not moderators of the obtained relationship.

There is a growing body of research literature which is supportive of the general hypothesis that performance is, in part, a function of the expectancies which others have of one's competency and ability. That is, to the extent that one is in a social, interpersonal environment in which others believe him to be competent and able and communicate such expectancies and beliefs by words and action, then, to that extent, performance is facilitated. These findings are, then, generally consistent with a general theoretical model which argues that performance is a function of one's self-perceived competence for the task at hand, with such self-perceived competence being, in turn, a function of one's previous experience with the task and the social definitions and reality which others provide for use concerning our competence for the task (Korman, 1970).

The purpose of this article is to add to this literature by reporting the results of a series of laboratory and field investigations which had as their general goal an attempt to see whether the fruitfulness of this hypothesis might be limited by various situational parameters of the type which are commonly found in organizational contexts. As will be seen, the conclusions from the five studies to be reported are all in the direction of not supporting any such parameter limitation, thus suggesting that the proposed hypothesis may have a wide degree of generality.

Laboratory Studies

One possible problem with the major body of previous research has been that it has been primarily concerned with "leader authority figure" expectancies of subordinate performance, that is, for the most part, the research has concerned itself with the effects of high-status individuals' expectancies as influences on the performance of low-status subordinates. Consider, for example, the studies of Berlew and Hall (1966) and Stedry and Kay (1966). In both of these field studies it was found that performance increased as a direct positive function of the level of expectation which their superiors in the organization had. Thus, in both cases, the more their superiors thought of them, that is, the higher their level of performance expectation as this was communicated to the subordinates, the higher the level of performance. While such research findings are of importance in and of themselves, it leaves open the question of whether such results would be obtained when the individuals having the expectancies did not have a formally defined superior role to those receiving the influence, but rather were in a peer relationship to them. Since such peer relationships may constitute a highly significant part of one's interaction pattern in many organizational situations, the purpose of the two experiments to be reported here was to test this general hypothesis when the expectancy of competency came from peers, rather than from authority figures.

In addition, a second parameter designed to test the generality of the hypothesis constituted the purpose of the second experiment.

---

[1] The author's appreciation is extended to P. Kaplan, B. Axel, and C. Lieutaud who contributed greatly to the various experiments and studies reported in this article.

[2] Requests for reprints should be sent to Abraham K. Korman, Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.

[3] Copies of the questionnaires used are available from the author upon request.

The parameter involved in this case was to see if the proposed hypothesis held up for those differing in levels of intelligence, the logic being that high intelligence individuals might be less likely to accept such influence than those lower on this variable.

Method

Subjects. The *S*s in both experiments consisted of college students fulfilling course requirements in introductory psychology. The actual sample sizes for the various treatments are indicated in Table 1.

Procedure: Exp. I. The procedure consisted of asking *S*s to respond to various creative tasks which were being studied by university psychologists as part of a general project designed to develop new types of intelligence tests which would not have the traditional weaknesses. Each group was also told that the same tasks had been previously taken by other college students and that these groups had been asked afterward to indicate what they considered to be an appropriate level of performance for college students on these tasks. The *E* would then list these performance expectancies for each task, with these differing for the high-expectancy group and the low-expectancy group. The actual levels used were chosen on the basis of previous research with these tasks with college students. The four tasks were (*a*) "how many uses can you think of for a brick?" (*b*) "How many things can you describe by the word 'sweet'?" (*c*) "How many different ways can you think of to attract tourists to the United States?" and (*d*) "How many different implications can you think of if a Federal law were passed providing for a free college education for everybody?"

Procedure: Exp. II. The procedure followed here was exactly the same as in Exp. I, except that *S*s were also asked to fill out the Wonderlic Personnel Test before responding to the creative tasks.

TABLE 1

EFFECTS OF PEER EXPECTANCIES ON PERFORMANCE

Task	High performance expectancy[a]		Low performance expectancy[b]		t
	$\bar{X}$	SD	$\bar{X}$	SD	
1	7.14	2.9	6.28	2.2	1.33
2	9.86	3.4	9.05	3.3	.97
3	6.77	2.4	4.83	2.0	3.04**
4	5.50	2.0	4.57	1.5	1.88*

Note.—All tests are one-tailed tests.
[a] $n = 22$.
[b] $n = 42$.
* $p < .05$.
** $p < .01$.

TABLE 2

EFFECTS OF PEER EXPECTANCIES ON PERFORMANCE FOR DIFFERING LEVELS OF SUBJECT INTELLIGENCE

Task	High performance expectancy[a]		Low performance expectancy[b]		t
	$\bar{X}$	SD	$\bar{X}$	SD	
			Total group		
1	7.85	2.6	5.90	1.8	4.79**
2	11.07	4.0	8.96	2.6	3.42**
3	6.93	3.4	5.14	2.2	3.45**
4	4.97	2.1	4.26	1.7	2.05*
			High intelligence		
1	7.37	2.4	6.03	2.1	2.37**
2	10.85	4.3	9.38	2.5	1.66*
3	6.78	3.3	5.23	2.5	2.11*
4	5.26	2.3	4.53	1.8	1.42
			Medium intelligence		
1	8.24	2.8	5.71	1.4	4.16**
2	11.24	3.7	8.37	2.5	3.22**
3	7.06	3.6	5.00	1.7	2.64**
4	4.73	1.9	3.87	1.6	1.80*

Note.—All tests are one-tailed tests.
[a] $ns = 60$, 27, and 33, respectively, for total group, high intelligence, and medium intelligence.
[b] $ns = 58$, 34, and 24, respectively, for total group, high intelligence, and medium intelligence.
* $p < .05$.
** $p < .01$.

Results

The results are indicated in Table 1 and 2 and provide, in general, overall strong support for the hypothesis. All results in Exp. I are in the direction predicted with two of four being significant, and all overall results in Exp. II are significant in the direction predicted. However, the higher intelligence does seem to decrease the effects of the expectancy manipulation slightly, only three of the four being significant in this case (one-tailed tests). The effect, however, does seem to be a powerful one, even for this group.

Field Studies

In the research to be reported here the general intention was to ascertain whether the hypothesis that "the higher the level of competence expected of employees, the higher the resulting work motivation and work performance" would be supported in different types of actual work contexts.

Method

Subjects. Study A consisted of a group of relatively low-level office employees in a Federal government office. What was of interest here, in addition to the level of the people involved, is that all of the people involved had civil service tenure, an employment condition that has been argued by some to lead to less responsiveness to job-related stimuli and less concern with job conditions. If this were so, one might predict that there should be a lessening of a relationship between leadership and other environmental variables and job motivation-performance.

Study B consisted of high-level technical employees in a data-processing center at a large university. All of the individuals involved here had at least one degree and a considerable number had completed either masters or doctoral study. Of considerable interest here was to see whether individuals of such relatively high abilities were susceptible to social influence effects of this nature in the "real world," as opposed to the laboratory findings of Study 2.

Study C consisted of a group of clerical employees in a branch of a large bank. This office was composed almost completely of high school graduates, and it was the opinion of the branch head that most had little promotion potential. In a very real sense this group is very similar to that in Study A, except that they were working for a private firm and were protected neither by civil service tenure nor by union regulations.

Procedure. 1. A questionnaire was developed which was designed to indicate the extent to which the employees felt they were being expected to engage in meaningful behavior and the extent to which high performance was expected of them. In addition, items were included that were to indicate the general level of the individuals' work motivation.

Items of the first type included such as (*a*) "when problems arise in my organization, the superiors generally try to get everybody's opinion as to how it can be resolved, in addition to their own"; (*b*) "the superiors believe that they do not have to watch you all the time in order to make sure that you are doing a good job"; and (*c*) "the management on my job seems to develop programs and policies which indicate they believe you are incapable of going any higher than you are in the organization." Items measuring self-reported work motivation included such as (*a*) "the likelihood that I would be more motivated to work if I were on a different job is very small"; (*b*) "most of the time I have to force myself to go to work"; and (*c*) "in terms of actual work motivation, compared to the people around here, I would place myself in the ____ile."

Identical forms of this questionnaire were utilized in Studies A and C and a slightly different version in Study B.

2. The questionnaires were administered in each case by people who were part of the work unit, but the project itself was identified as a university project.

3. Performance evaluations were obtained from relevant superiors for Studies A and B but were not obtainable in Study C. Hence, the criterion for the latter study was self-reported work motivation as measured by four questionnaire items. Some evidence for the construct validity of this latter measure is that it was correlated .93 with performance evaluations for the Study A sample.

4. The analysis of the data in all cases consisted of dividing the criterion variable into those considered "excellent" and "average" and then computing mean differences for each of these groups on each of the items measuring the degree to which the leadership and managerial system was perceived as having high expectations of competency and performance. Sign tests were then computed in each study to test whether the criterion groups had consistently different perceptions of leadership and managerial expectations.

Results

The results were as predicted in all cases. For Study A, the "high" group ($n = 9$) scored higher than the low group ($n = 22$) on 11 of the 14 items with one tie ($p < .05$); for Study B the comparable figures were 9 of 11 ($p < .05$) (n of "highs" $= 6$ and n of "lows" $= 22$); while the "high" group ($n = 12$) was higher than the "low" group ($n = 16$) on all 14 items in Study C.

DISCUSSION

The general findings of these studies that expectations from others are related to performance, regardless of the source of the expectancy or the characteristics of the receiver of the expectancy, has a number of interesting theoretical and practical implications.

The implications for social organizations seem clear in that the need for a social reality to define oneself and the desire to perform consistently with such reality is not limited to those situations where clear, formally defined authority figures are available, or just to people of low intelligence and/or previous achievement. Apparently, we use whatever we can, authority figure or not, intelligent or not, to define ourselves vis-à-vis a given situation, and then we behave in a way consistent with such a definition. Thus, one interesting implication of these findings is that high expectancies can be established in an organization by working through the medium of work group norms and group values, as well as through the medium of leadership influence directly. This opens the door for a greater variety of ways by which the motivation to work may be increased by organizational leadership.

One question which arises here is that while our suggested mechanism for work improvement might not be disputed, our theoretical interpretation might be. For example, it might be argued that the influence of peer expectancies as we have studied them here has an influence not on self-conceptions but, rather, on the conscious goals which one is striving for. That is, the higher the expectancies, the higher the level of conscious goal-setting and the higher the performance. This type of theorizing, particularly exemplified in the work of Locke (1968), suggests a much more rational model of man than the interpretation we have advanced here. Our preference is for the latter, however, on the ground that it is able to account both for the positive findings of the rational approach and for those research results which have negative implications for this type of theorizing. The way in which we can do this is by differentiating between externally and internally set goals.

Taking the latter case first, we would argue that the propensity to achieve high self-set goals would be a linear function of self-esteem since such achievement would be consistent with high self-esteem. On the other hand, people of low self-esteem who have high self-set goals should not be motivated to achieve them. These predictions have been supported in previous research by the writer (Korman, 1970), Zander and Newcomb (1967), and in a study by Bryan and Locke (1967). In the latter study it was found that those who had a high opinion placed on them worked harder to achieve the same self-set goals than those who had a low opinion placed on them but had the same self-set goals.

On the other hand, when the goals are externally set, both the consistency model and the rational model would predict that performance is a function of the level of expectation, a prediction which is well supported by the findings of Berlew and Hall (1966), Stedry and Kay (1966), and the research reported in this article.

Another implication of these findings concerns contemporary leadership theory. In the continuing controversies which have marked the growth of modern or behavioral science organization theory as opposed to the more traditional frameworks, there has been a recent trend among some students of the field suggesting that they may both be right. That is, perhaps one mode of leadership is correct for one type of individual or work group and the leader should adjust his behavior to be appropriate to the characteristics of that group and that, in a different situation and with a different group, perhaps another form of leadership might be most appropriate and most beneficial.

While this point of view has a great degree of appeal, there are some serious problems which develop from its adoption. One of these is that such a theoretical stance seems to project a picture of a leader as a "reactor" to the group and a person who fits himself to the group, rather than as an "innovator" and as an instrument for "change" and "improvement." Second, while a leader may adapt himself to a group and fit his style to their characteristics, given the group as it is, such adaptation ignores the simple observation that perhaps the best way to increase the group's effectiveness would not be to adapt the leader's behavior to the group, but to change the group. Finally, a third problem with such attempts at theoretical reconciliation is that they may not be necessary in that a specific mode of leadership behavior may have more wide-ranging applicability than previously thought. The findings of these studies suggest considerable support of this third position (without negating in the slightest the fact that the other two factors may also have considerable importance).

There are, of course, some limitations to the research reported here, with the two most important being (*a*) the sample size reported in the field studies; and (*b*) the fact that these latter studies were correlational, thus suggesting some ambiguity in interpretation. Our response to these limitations would be as follows. First, while there is little doubt that high expectancies might result from high performance as well as the other way around, as we ourselves have argued elsewhere (cf. Korman, 1966), this interpretation cannot account for the findings of the laboratory studies. On the other hand, the interpretation we have advanced will account for the findings of all the studies.

Second, while there is little doubt that the sample sizes in the last three studies are quite small when viewed separately, the studies are highly consistent in their findings. Similarly, if we had standardized each distribution and then pooled them, then the samples would have more closely approached respectability, although still on the small side. However, we choose to present them this way in order to be able to show the consistency of findings across different types of work situations. Hence, while further replication of these findings with larger samples is greatly needed, the consistency of the obtained findings and their integration with the first two studies should also be kept in mind.

One final note concerning needed further research concerns how far we can reasonably push the expectancy effect. It is clear that an asymptote will be reached, but where? How high? Can we formulate some general law? For example, might it be some ratio of the level of performance a person will engage in when he is under little or no external expectancy? Or will the general law, if there is one, depend on the original level of performance?

REFERENCES

BERLEW, D. E., & HALL, D. T. The socialization of managers: Effects of expectations on performance. *Administrative Science Quarterly*, 1966, **11**, 207–223.

BRYAN, J., & LOCKE, E. Parkinson's law as a goal-setting phenomena. *Organizational Behavior and Human Performance*, 1967, **2**, 258–275.

KORMAN, A. K. Consideration, initiating structure and organizational criteria—A review. *Personnel Psychology*, 1966, **19**, 349–362.

KORMAN, A. K. Toward a hypothesis of work behavior. *Journal of Applied Psychology*, 1970, **54**, 31–41.

LOCKE, E. A. Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 1968, **3**, 157–189.

STEDRY, A. C., & KAY, E. The effects of goal difficulty on performance. *Behavioral Science*, 1966, **11**, 459–470.

ZANDER, A., & NEWCOMB, T., JR. Group levels of aspiration in United Fund campaigns. *Journal of Personality and Social Psychology*, 1967, **6**, 157–162.

(Received February 18, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 223–225

DEVELOPMENT OF A MEASURE OF UNEMPLOYABILITY AMONG RESTORED PSYCHIATRIC PATIENTS [1]

DONALD J. SEARLS,[2] LOWELL T. WILSON, AND R. W. MISKIMINS
Colorado State University

The vocational rehabilitation of the mentally restored patient is an area in need of a practical strategy to aid in identifying those who would have difficulty in making a satisfactory vocational adjustment. This article examines the tenability of supplementing subjective clinical information concerning a patient's employability with more objective data for formulating specific placement procedures. The analysis revealed that 15 behavioral items differentiated placed successes, placed failures, and nonplaced patients on an initial and replication analysis. Validation and replication results indicated that the final items comprise an effective instrument for selecting the mentally restored individual who will experience failure in employment.

The vocational counselor in the mental hospital setting must strike an optimum balance in time expenditure between patients who will clearly benefit from his efforts and the unemployable client who might better benefit from some other hospital procedure. Continued reliance upon clinical judgment as to a patient's employability is not only time-consuming, but dilutes the potentiality for deriving full advantage from placement services.

The objective of this study was to examine the characteristics of patients classified as hard-core unemployables and to develop a methodological approach to predicting vocational failure. This investigation was undertaken to determine the feasibility of supplementing the more subjective clinical information concerning a patient's employability with more objective data for the purpose of formulating specific placement procedures.

Past research shows that predictive studies in the area of work adjustment have taken a more positive approach; that is, predicting the potential vocational success. Several studies have examined variables important to job success (e.g., Brown, 1966; Distefano, 1966; Wilson, Miskimins, & Berry, 1969) and job satisfaction (Robinson, Connors, & Robinson, 1964). However, there are no investigations representative of the opposite approach; that is, developing criteria by which to identify patients most unlikely to make a successful vocational adjustment.

The setting for this investigation was the Fort Logan Mental Health Center (Mental Health and Manpower Project, 1968) in Denver, Colorado. As an integral component of the Center's vocational services, a placement department operates to develop jobs in the community and to place restored psychiatric patients in appropriate positions. Information collected by this department was the primary source of data relevant to the unemployable patient.

METHOD AND RESULTS

Phase 1

The placement department began by reviewing its records of patient-clients who had been both successfully and unsuccessfully placed and employed. A pool of items was then compiled listing behaviors incompatible with successful employment. The items consisted of both preemployment (i.e., interview situations) and on-the-job behaviors.

After screening the items to minimize ambiguity and item overlap, a total of 40 behavioral variables were selected for analysis. These items were then classified within the following categories: (*a*) motivational: for example, "late for placement interview"; (*b*) employment experience and stability: for example, "no experience"; (*c*) vocational goal: for example, "unrealistic in terms of training

---

[1] This investigation was supported in part by the United States Department of Labor (OMPER Grant No. 82-06-66-62).

This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Lowell T. Wilson, Manpower Studies, Colorado State University, 50 West Fifth Avenue, Denver, Colorado 80204.

TABLE 1

ANALYSIS OF ITEMS INCOMPATIBLE WITH VOCATIONAL SUCCESS

Item	χ^2
Motivational	
Misses appointments for placement or counseling interviews	12.34*****
Usually late for placement or counseling interviews	4.46*
Resistance or refusal to accept job referrals	7.65***
Procrastination on following-up job leads	12.70*****
Vocational	
No work experience	3.83*
Vocational goal too limited	3.65*
Vocational goal unrealistic	3.56*
Incomplete or no vocational or technical training, and no education beyond high school	8.99****
No transportation, or dependent on public transportation which is unreliable or difficult to use	3.52*
Personal	
Inappropriate behavior during placement or counseling interview	4.61**
Extreme under- or overweight	4.35*
Appearance of having a lack of physical vitality	9.15****
Limited because of physical handicap	3.56*
Difficulty in making self understood (poor communication of thought)	4.96**
No evidence of being able to handle leadership responsibilities	7.73***
Indicates preference to being a follower rather than a leader	8.66****

* $p \leq .10$.
** $p \leq .05$.
*** $p \leq .025$.
**** $p \leq .01$.
***** $p \leq .001$.

and experience"; (*d*) personal appearance: for example, "extremely unkempt"; (*e*) personality: for example, "antagonistic"; (*f*) mental alertness, self expression: for example, "apathetic"; (*g*) miscellaneous: for example, "too old."

Phase 2

After selection of items, a group of 87 patients was drawn, all of whom had at one time during hospitalization received service from the placement department. This group was then divided into three categories of vocational success defined by the following criteria: (*a*) placed failures (PF): patients who had terminated within 3 mo. after placement ($N = 25$), (*b*) placed successess (PS): patients still employed 3 mo. after placement ($N = 32$), (*c*) nonplaced patients (NP): patients who had utilized placement services, but were not placed ($N = 30$). The three groups were then statistically compared on the 40 behavioral items selected in Phase 1. The initial analysis indicated that 16 of the variables either tended to differentiate or differed significantly among the three groups. These items were classified into the categories of motivational, vocational, and personal characteristics. Table 1 summarizes these findings.

Phase 3

The 16 items retained from the foregoing analysis were then cast into the form of a checklist. It was presumed that the scale was an effective instrument in differentiating among the three vocational groups. However, in an effort to increase the reliability and sensitivity of the present instrument, a replication analysis was performed utilizing the items that showed differences in Phase 2.

A new sample of patients was selected consisting of the following groups for the replication analysis: (*a*) PF ($N = 41$), (*b*) PS ($N = 49$), (*c*) NP ($N = 39$). This group of patient-clients was then statistically compared on the 16 items found to differentiate the validation groups in Phase 2. The replication analysis demonstrated that 15 of the items again produced meaningful differences. The final listing of items comprising the Hard-Core Unemployable Scale (HUS) and validation data are presented in the following section.

Phase 4

The validation and replication (not a true cross-validation due to unavailability of *S* data) of the HUS took the form of rating each of the original groups on the 15 items constituting the scale. Counselors were asked to rate each patient in the two groups (validation group, $n = 87$, replication group, $n = 129$) on the items incompatible with successful vocational adjustment. The mean differences for the groups are summarized in Table 2. The mean values in Table 2 indicate that the scale is effective in selecting the three groups of patients as distinct groups according to degree of vocational success.

TABLE 2

SUMMARY OF MEAN DIFFERENCES PRODUCED ON HARD-CORE UNEMPLOYABILITY SCALE RATINGS

Group	$\bar{X}$	F
Validation		22.32*
NP ($n = 30$)	12.77	
PF ($n = 25$)	10.08	
PS ($n = 32$)	4.19	
Replication		19.62*
NP ($n = 41$)	12.34	
PF ($n = 39$)	9.95	
PS ($n = 49$)	4.12	

Note.—Abbreviations: NP = nonplaced patients, PF = placed failures, and PS = placed successes.
* $p = .001$.

DISCUSSION

The final 15 items were classified as motivational, vocational, or of a personal nature. The hardcore unemployable patient-client is definitely lacking in motivation for the placement process. This person is prone to missing counseling interviews, reluctant in accepting job referrals, and procrastinates in following-up on potential employment leads.

The unemployable person has a poor work history. Typically, the vocational goal is much too narrow, unrealistic in terms of his abilities, or is uncertain. This individual also creates a negative impression during the interview situation.

The most consequential implication of identifying such behaviors as a part of the present strategy is the illustration of the potential for employing objective data to supplement more clinically oriented information concerning a patient's employability. Dependence on clinical judgment alone can result in notable wastes of counselor time and energy when dealing with the chronically unemployable patient-client. The practical approach offered by this research can serve as an aid in better distributing counselor time according to client needs.

Examination of the item content reveals the specific reasons or trouble areas for a particular unemployable client. For example, a client may be rated as potentially unemployable due partly to missing employment interviews or procrastination in following-up on job leads. This immediately informs the counselor of specific behaviors detrimental to this person's successful employment. Intensified counseling and retraining by the vocational counselor could increase the probability of these individuals obtaining successful employment.

Subsequent validation with larger groups should produce clearer and more accurate normative data. Also, inspection of the content of items indicates that they are not necessarily behaviors that are characteristic of mental illness. Many unemployable people are characterized by qualities such as "no work experience" or "lack of education." Therefore, the items appear to envelop and are applicable to the whole range of unemployable people, including the mentally disturbed client.

REFERENCES

BROWN, R. L. Imperatives for job success. *Personnel Journal*, 1966, **45**, 478–483.

DISTEFANO, M. K., PRYER, M. W., & RICE, D. P. Orientation and job success of mental hospital patients. *Psychological Reports*, 1966, **19**, 113–114.

MENTAL HEALTH AND MANPOWER PROJECT. Denver: Fort Logan Mental Health Center, 1968.

ROBINSON, H. A., CONNORS, R. P., & ROBINSON, A. H. Job satisfaction researches of 1963. *Personnel and Guidance Journal*, 1964, **43**, 360–366.

WILSON, L. T., MISKIMINS, R. W., & BERRY, K. L. An assessment of characteristics related to vocational success among restored psychiatric patients. *Vocational Guidance Quarterly*, 1969, **5**, 110–114.

(Received December 22, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 226–228

SOURCES OF REFERRAL AND EMPLOYEE TURNOVER[1]

MARTIN J. GANNON[2]

University of Maryland

In this study seven methods or sources that are used to obtain new workers are compared in terms of their influence on employee turnover. Four sources are shown to be predictors of stable employment, namely, the reemployment of former workers who left but now desire to return, the hiring of individuals referred by their high schools, the hiring of individuals referred by present employees, and others (primarily walk-ins). Conversely, three sources are associated with high employee turnover, namely, the utilization of hiring agencies other than the major one under contract with the firm, newspaper advertising, and the use of the major hiring agency.

One of the presistent theoretical and practical problems facing industrial psychologists is employee turnover. The literature on this problem stresses a variety of predictors such as age, relations with supervisors, and interest tests (Schuh, 1967; Vroom, 1964). One possible explanation of employee turnover may be the sources from which employees are drawn. In this regard, previous published research has primarily delineated sources of referral or recruiting patterns without relating them to employee turnover. For example, Mandell (1956) shows that companies with 500 or fewer employees tend to favor employment agencies while larger firms generally have recourse to newspaper advertisements (Mandell, 1956; Malm, 1954).

The objective of this article is to test the hypothesis that recruiting patterns or sources of referral are predictive of employee turnover. If the hypothesis is valid, it should be possible to eliminate unstable employees even before the application of interest and psychological tests through the use of particular sources of referral.

METHOD

The personnel department of a New York bank has been studying the problem of employee referral for several years; the data presented in this article have been supplied by this bank. This department categorizes their seven sources of referral as follows: (*a*) The employee saw an advertisement for work in a newspaper. (*b*) The employee was hired by the major hiring agency under contract with the bank. (*c*) The employee was hired by hiring agencies other than the major one under contract with the bank. (*d*) The employee was referred to the bank by a present employee. (*e*) The employee has been reemployed after a period of time away from the job. (*f*) The employee was referred by his high school. and (*g*) The employee came to the bank through some other method, mainly by walking into the personnel office and asking for work. For each source of referral, the bank then develops a quit rate, that is, the number of employees hired through a particular source within a specific year divided into the number of employees from that source who were hired but quit within that year. Thus, the quit rates can be compared in order to discover the best and worst sources of referral.

In order to ascertain if the differences in the quit rates were significant, 2×2 tables were constructed, the rows of which were two sources of referral and the columns of which were the number of employees not quitting and the number who quit. Chi-square was then used to test the significance of the differences (see Table 2).

RESULTS

Four years are investigated: 1961–1964. In Table 1, the quit rates for each of the seven sources of referral are presented. A total of 6,390 employees were hired during this period, and 1,934 terminated before completing their first year of service with the bank.

As shown in Table 1, the quit rates range from a low of 21.3% for employees who were reemployed to a high of 40.0% for employees who came to the bank through the major hiring agency. Table 2 is a matrix of chi-square values showing the association between the sources of referral. Clear and significant differences between the sources of referral were discovered. Thus, four sources of referral were uniformly and significantly different from three other sources of referral ($p \leq .001$). The four sources are the reemployment of former workers of the

---

[1] The author would like to thank John B. Miner and Stephen J. Carroll for helpful criticisms and suggestions.

This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to the author, Behavioral Science Division, College of Business and Public Administration, University of Maryland, College Park, Maryland 20742.

bank, the hiring of individuals referred by their high schools, the hiring of individuals referred by present employees, and others (primarily walk-ins). These four sources of referral have low quit rates ranging from 21.3% to 28.8% and can be considered as suppliers of stable employees. Conversely, the three sources of referral which have quit rates ranging from 37.8% to 40.0% can be viewed as suppliers of unstable employees. These three sources are the employment of individuals who come to the bank through the major hiring agency, the employment of individuals who come because of newspaper advertising, and the employment of individuals who are contacted through hiring agencies other than the major one.

There are no significant within-group differences for the three poor sources of referral. However, within the group of good sources there are three significant differences. "Others, mainly walk-ins" is a poorer method for recruiting employees than reemployment ($p \leq .05$) or recommendations by high school administrators ($p \leq .01$), although it does not significantly differ from recommendations by fellow employees. Reference by present employees is a poorer source than recommendations by high school administrators ($p \leq .05$).

TABLE 1

PERCENTAGE QUIT RATE BY SOURCE OF REFERRAL: 1961–1964

Source of referral	No. hired	No. quit	% quit
Reemployed	253	54	21.3
Referred by high school	602	131	21.8
Referred by present employees	2320	615	26.5
Others (mainly walk-ins)	1212	349	28.8
Hiring agencies except "major"	637	241	37.8
Newspaper advertising	512	202	39.4
Major hiring agency	854	342	40.0
Total	6390	1934	

The final methodological point of importance concerns the stability of quit rates for each method over time. In the analysis thus far the total figures for the period 1961–1964 have been presented. It is entirely possible that a particular recruiting method is characterized by a low quit rate one year but a high quit rate the next. Such inconsistencies would tend to undermine the analysis. However, as shown in Table 3, there are no inconsistencies when the

TABLE 2

MATRIX OF CHI-SQUARE VALUES SHOWING THE ASSOCIATION BETWEEN THE SOURCES OF REFERRAL

Source of referral	Lower quit rates[a]				Higher quit rates[b]		
	Reemployed	Referred by high school	Referred by present employees	Others (mainly walk-ins)	Hiring agencies except "major"	Newspaper advertising	Major hiring agency
Reemployed		.04	3.40	6.14***	22.83*	25.57*	30.46*
Referred by high school			5.67***	10.23**	38.06*	41.33*	53.83*
Referred by present employees				2.10	31.16*	34.24*	54.32*
Others (mainly walk-ins)					15.70*	18.80*	28.49*
Hiring agencies except "major"						.31	.75
Newspaper advertising							.05
Major hiring agency							

[a] 21.3%–28.8%.
[b] 37.8%–40.0%.
* $p \leq .001$.
** $p \leq .01$.
*** $p \leq .05$.

TABLE 3

CONVERSION OF THE QUIT RATES FOR EACH RECRUITING METHOD TO RANKS FOR EACH YEAR

Recruiting method	Ranked data			
	1961	1962	1963	1964
Newspapers	7	6	5	5
Major hiring agency	6	7	7	—[a]
Other agencies	5	5	6	—[a]
Referred by fellow employee	3	2	2	3
Reemployed	1	1	3	2
Referred by high school	2	4	1	1
Others (mainly walk-ins)	4	3	4	4

[a] No employees were hired through these sources in 1964.

quit rates are analyzed by year. In this table the quit rates for each year have been converted to ranks. In each year the three poor methods of recruitment uniformly have higher quit rates than the four good methods of recruitment. While there are some differences within groups—for example, newspapers ranked seventh in 1961 but sixth in 1962—there are no exceptions when the two groups of good and poor recruiting methods are compared.

DISCUSSION

Even before an employee approaches a firm in order to fill out the application form, it is possible to improve the selection process by utilizing some sources of referral rather than others. This study has demonstrated that four sources of referral are predictive of stable employees, namely, the reemployment of former workers of the bank, the hiring of individuals referred by their high schools, the hiring of individuals referred by present employees, and others (primarily walk-ins). However, three sources of referral are predictive of unstable employees, namely, the use of hiring agencies other than the major one, newspaper advertisements, and the major hiring agency.

The practical significance of this study is relevant for many firms. There has been some research into the cost of hiring and training new employees (see Canfield, 1959, Gaudet, 1958; and McManemin, 1960). A conservative estimate of such costs is $1,000 per employee. The significant difference in the quit rates for the seven sources is 9.0%, since the two groups of sources identified as good and poor in reference to employee turnover are divided into their respective components by "others (mainly walk-ins)" (28.8%) and "hiring agencies other than the major one" (37.8%). In the 4 yr. under examination, approximately 2,000 employees left the bank without completing their first year of employment. Given a quit rate of nine per 100 that could have been avoided if the three poor sources had not been used, the total cost to the bank is $180,000 (9 × 20 × $1,000).

This analysis assumes that enough employees could be gathered through the four good sources of referral. In a tight labor market, such an assumption may be unrealistic. However, the analysis does suggest that, even if a firm were to spend $40,000 or $50,000 in increasing the number of individuals hired through the four good sources rather than through the three poor sources, the investment would probably be a good one.

Although further research into the sources of referral is necessary so that the demographic and attitudinal characteristics of these workers can be assessed, the present analysis indicates that a substantial proportion of employee turnover can be eliminated through the judicious use of various procedures for attracting employees.

REFERENCES

CANFIELD, G. How to compute your labor turnover costs. *Personnel Journal*, 1959, **37**, 413–417.

GAUDET, F. What top management doesn't know about turnover. *Personnel*, 1958, **34**, 54–59.

MALM, F. Recruiting patterns and the functioning of labor markets. *Industrial and Labor Relations Review*, 1954, **7**, 507–525.

MANDELL, M. *Recruiting and selecting office employees*. New York: American Management Association, 1956.

MCMANEMIN, J. A practical method of calculating turnover cost. *Personnel*, 1960, **37**, 73–77.

SCHUH, A. The predictability of employee tenure: A review of the literature. *Personnel Psychology*, 1967, **20**, 133–152.

VROOM, V. *Work and motivation*. New York: Wiley, 1964.

(Received December 19, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 229–233

PERSONNEL SELECTION AND THE PREDICTION OF ERROR[1]

I. VAN W. RAUBENHEIMER[2] AND JOSEPH TIFFIN

University of South Africa at Pretoria *Purdue University*

A few trends in current personnel selection research are summarized to constitute a backdrop against which the rationale for a prediction model, taking advantage of the concept of underpredictability and overpredictability, is being presented. A noncognitive predictor of predictability is developed ($n = 224$) and used to subgroup a cross-validation sample into underpredictables ($n = 25$), accurately predictables ($n = 51$), and overpredictables ($n = 25$). Product-moment correlations between the unweighted sum of six cognitive predictors (subtests on the Purdue Clerical Adaptability Test) and a criterion measure of overall job performance resulted in $r_{xy} = .26$, $r_{xy} = .48$, and $r_{xy} = .40$ for the three subgroups, respectively. This compared favorably with an $r_{xy} = .25$ before subgrouping and an $r_{\hat{y}y} = .34$, involving the above six cognitive predictors and two noncognitive predictor scales (used to develop the predictor of predictability).

Over the last few years some interesting articles appeared on various off-quadrant topics such as differential predictability (Ghiselli, 1956, 1960a, 1960b, 1963; Ghiselli & Sanders, 1967), moderator variables (Saunders, 1956), the twisted-pear phenomenon (Fisher, 1959), intervening variables (Dunnette, 1963), and moderator designs (Guion, 1967).

In addition to reflecting a growing uneasiness about what has been called the "traditional prediction model" all of these articles call for a fresh and probably more complex but realistic approach toward prediction problems. By implication they also repudiate the applicability of well-established psychometric theories adhered to in this field.

Although some authors (e.g., Storms, 1960; Abrahams, 1969) take issue with some of the ideas expressed in the aforementioned articles, the deviation from traditional or conventional prediction strategies seems to have some general if not rational appeal.

The analyses and prediction model to be proposed in the present article were developed primarily against some current research on the concept of predictive error and its relation to specific individuals in specific prediction situations.

It was argued that if error of prediction were systematic and therefore if it appears to recur in one situation (or group) after another, it should be possible to develop a predictor of such error.

According to Ghiselli, error of prediction is defined in terms of the difference (D) between the predicted and observed criterion scores for a given group of *S*s. Although Ghiselli did not attach any signs to this difference, it is suggested here that the algebraic difference between the observed and predicted criterion scores would make for a finer analysis, since it will not only indicate the magnitude of predictive error, but will also tell what the direction (overpredicted or underpredicted) of the predictive error for each *S* would be. Furthermore, the difference scores would seem to be predictable by something like a noncognitive predictor scale (P) consisting of biographical items validated against the difference scores as the criterion.

If the aforementioned predictor (P) could be developed, it could be used to estimate the extent to which each individual in a specific situation tends to be overpredicted or underpredicted by a predictor (X) or battery of predictors already in use.

In addition, it is suggested that any group of individuals in a given situation actually consists of three relatively independent subgroups—(*a*) those who are relatively accurately predictable by a given battery of predictors, (*b*) those that are relatively overpredictable, and (*c*) those who tend to be relatively underpredictable.

The predictor of error (P) as discussed above, can be used to define these subgroups. This can easily be done since the predictor of error estimates for each individual the magnitude as well as the direction of predictive error involved. Thus, it could be inferred that all individuals who score above the mean on the predictor of error (P) tend to be overpredicted and all those that score below the mean tend to be underpredicted. This is the specific advantage of obtaining the algebraic and not the absolute difference between the predicted and observed criterion scores.

---

[1] This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to I. van W. Raubenheimer, Department of Industrial Psychology, University of South Africa, P.O. Box 392, Pretoria, South Africa.

The subgrouping itself would be rather arbitrary. One can, for example, regard the overpredictables as the 25% of the group that score highest on P, and the 25% that score lowest on P can be considered underpredictable. This will leave the middle 50% of the group defined as predictable. It can readily be seen that these cut-off points could be altered depending on the number of employees to be selected, the number of applicants available, etc.

Once the subgroups are defined the original battery of cognitive predictors can be applied to each group separately. It is expected that the correlation between the predictors and the criterion will be higher for each subgroup separately than for the total group taken together.

METHOD

Subjects. The total sample used in this study consisted of 339 male and female Afrikaans-speaking clerical workers employed by a large commercial bank in South Africa. The bank operates through branches established at different locations throughout a large city. The *S*s thus came from different branches though the work performed at the different locations appeared to be quite comparable.

All material used in the investigation was presented in English. This, however, was not regarded as a major obstacle since all employees of organizations providing public services in South Africa are required to be bilingual (Afrikaans-English). In fact, the South African educational system provides for double medium instruction at all levels. It could therefore be expected that any extent of difficulty imposed on *S*s by using English material would be more or less the same for all.

Work performed. According to a brief outline of the organizational structure and type of operations performed, as furnished by the head manager of personnel, the daily activities engaged in by *S*s can best be described as being of a general clerical nature. This description seems to be particularly appropriate since it has been a policy of the bank to rotate workers periodically as to induce substitutability and flexibility of its personnel.

Criterion of job success. Ratings of overall job success were secured by using a system of paired comparisons. The personnel force of each branch operates as a number of smaller groups with one person (the supervisor) in each group responsible for the performance of the members of his group. The supervisors in all the different branches were requested to act as judges of their employees' overall job success by using a paired-comparison rating system (Tiffin & McCormick, 1965).

All criterion ratings were gathered before any predictor information was obtained to reduce the probability that performance on the predictors could influence the criterial ratings made by the supervisors.

Predictors

Cognitive predictors. The Purdue Clerical Adaptability Test was administered to all employees for whom criterion data were obtained. The test, covering a wide range of clerical activities such as spelling, computation, checking, word meaning, copying, and reasoning was considered to provide meaningful information for the purposes of this study.

Although in the computation and reasoning subtests, reference to the dollar is frequently made, this is considered to have created no substantial problem since South Africa's currency (the rand) also operates on the decimal system.

Noncognitive (biographical) predictors. A sample of 100 biographical items was drawn from *A Catalogue of Life History Items* (Glennon, Albright, & Owens, 1966) with the number of response options per item ranging from two to eight. The content areas covered were habits and attitudes, health, human relations, money, family life (past and present), recreation, hobbies, interests, school, education, self-impressions, values, opinions, preferences, work history, and some classification data.

Response categories of some of the items clearly constituted a response continuum. In the case of several others, however, the assumption of an existing underlying response continuum could hardly be made. It was decided therefore, to regard each response option as being an individual biographical item with an assigned value of one if the response option was selected and zero if not. The procedure generated a total number of 494 binary scored items which provided a wealth of biographical information from which biographical predictor scales could be developed, as will be discussed.

Accuracy of Some Conventional Prediction Techniques

The following analyses were deemed necessary to secure some yardstick against which any improvement on prediction in this study could be evaluated.

The predictive effectiveness of the six cognitive predictors was determined by developing regression weights on an experimental group of 224 *S*s. The weights were then applied to a cross-validation group ($n = 110$) and the correlation between the observed and predicted criterion scores computed.

Two scales were developed to determine the accuracy of prediction of biographical (or noncognitive) predictors in this study. The point-biserial correlations between 494 binary items and the criterion were generated and two scales were derived. One consisting of 12 items which correlated positively with the criterion (Y) and the other of 20 negatively correlated items. All correlations were significant beyond the .01 level. Scores on the two noncognitive scales were obtained for 224 *S*s in an experimental group. After regression weights were developed for the two scales on the experimental group, they were applied to a cross-validation group consisting of 110 *S*s, and the correlation between the observed and predicted criterion scores could be determined.

In order to get an indication of the predictive accuracy of the six cognitive and two noncognitive predictors combined, a regression analysis was performed for the eight predictors on the experimental group ($n = 224$) and the weights so established were applied to a cross-validation group ($n = 110$).

Improvement of Prediction

In order to define the magnitude and direction of predictive error of the cognitive predictors, the algebraic differences between the standardized total scores on the Purdue Clerical Adaptability Test and the standardized criterion scores of all individuals in an experimental group ($n = 224$) were obtained. These algebraic difference scores constituted the criterion with which all 494 binary-scored biographical items were then correlated. Only those items ($n = 41$) with point-biserial correlation coefficients significant beyond the .05 level were selected and combined into a predictor of error scale.

A total score for each individual in the experimental group was obtained by assigning a + 1 if the individual endorsed an item which happened to be positively correlated with the difference score (D) and a − 1 whenever a negatively correlated item was endorsed. (This scoring system yielded 224 predictor of error scale scores ranging from + 10 to − 10—a positive score indicating a tendency to be overpredictable and a negative score just the reverse.)

The predictor of error (P) was then correlated with the criterion of error (D) in order to obtain some indication of its validity. It was realized, however, that the real test of P would come only under cross-validation, and it was therefore decided to apply the scale to a cross-validation group consisting of 101 *S*s. (It will be noticed that the sizes of the cross-validation groups used in the different situations were not always the same. In some instances cases had to be deleted since one of the computer programs used read blanks on the input cards as missing data instead of zeros as it was intended to be.)

In order to establish the three predictability subgroups as defined before, all individuals ($n = 25$) in a cross-validation group of 101 who scored above + 2 on the predictor of error scale were considered overpredictables, all those ($n = 25$) who scored below − 3 on P were regarded as underpredictables, and the rest ($n = 51$) constituted the accurately predictable group. The unweighted sum of the six cognitive predictors (X) was used as the predictor in this analysis, and the correlation between X and Y was obtained for each subgroup separately and combined to determine whether the subgrouping had any effect on the accuracy of prediction.

TABLE 1

MULTIPLE CORRELATION COEFFICIENTS AND THEIR CORRESPONDING CROSS-VALIDATION CORRELATION COEFFICIENTS FOR SIX COGNITIVE AND TWO NONCOGNITIVE PREDICTORS SEPARATELY AND COMBINED

Predictor	Experimental group $(R)^a$	Cross-validation group $(r_{yy})^b$
6 Cognitive (A)	.30	.27
2 Noncognitive (B)	.45	.33
A + B	.50	.34

[a] $n = 224$ in each predictor.
[b] $n = 110$ in each predictor.

RESULTS AND DISCUSSION

Accuracy of the Conventional Prediction Techniques

The multiple *R*s obtained for the six cognitive predictors and the two noncognitive predictor scales separately and combined appear in Table 1 along with their magnitudes under cross-validation. All *R*s were expected to shrink under cross-validation to some extent. There was, however, more shrinkage for the two noncognitive than for the cognitive predictors. Also the multiple correlation coefficients based on both cognitive and noncognitive predictors decreased more under cross-validation than the cognitive predictors alone.

The two noncognitive predictor scales seem to have done substantially better than the six cognitive predictors ($R = .45$ vs. $R = .30$) but were less stable under cross-validation ($r_{\hat{y}y} = .33$ vs. $r_{\hat{y}y} = .27$). Combined into a regression equation the eight predictors yielded an $R = .50$ which shrank to $r_{\hat{y}y} = .34$ under cross-validation. The latter correlation coefficient seemed to be the highest that could be obtained through an ordinary combination of the cognitive and noncognitive predictors into a regression equation. Any improvement in predictive accuracy therefore had to be compared with this figure.

Improvement of Prediction

Table 2 includes several correlation coefficients to indicate the extent of predictive accuracy under the different analyses proposed in the foregoing sections.

Predictor of error. A product-moment correlation of $r_{xy} = .57$ between the predictor of error

TABLE 2

Product-Moment Correlation Coefficients Indicating the Predictive Accuracy of Different Models of Prediction

Model	Predictors involved	Coefficient for the experimental group	Coefficient for the cross-validation group
Multiple regression	6 Cognitive and 2 noncognitive predictors	$R = .50^{a}$	$r_{\hat{y}y} = .34^{b}$
D-score	Predictor of error	$r_{PD} = .57^{a}$	$r_{PD} = .22^{c}$
Total group	Total on the 6 cognitive predictors		$r_{xy} = .25^{c}$
Underpredictable	Total on the 6 cognitive predictors		$r_{xy} = .26^{d}$
Overpredictable	Total on the 6 cognitive predictors		$r_{xy} = .40^{d}$
Accurately predictable	Total on the 6 cognitive predictors		$r_{xy} = .48^{e}$

Note.—All coefficients significant at $p < .01$. See text for explanation of abbreviations.

[a] $n = 224$.
[b] $n = 110$.
[c] $n = 101$.
[d] $n = 25$.
[e] $n = 51$.

and the criterion of error for the experimental group, and one of $r_{xy} = .22$ for the cross-validation group clearly indicates (although subject to considerable shrinkage) that it had been possible to develop a predictor of error in this study.

If this outcome is to be interpreted with reference to the individuals (as opposed to the predictors) involved in the prediction process it would be reasonable to infer that some individuals tended to interact differently with the six cognitive predictors than others. This differential interaction tendency did not only occur in the experimental group (which would indicate that it could be a random phenomenon), but it also proved to be predictable in a cross-validation sample.

A word about the concept of unpredictability seems in order at this juncture. Unpredictability, as indicated before, is usually defined in terms of the absolute difference between predictor and criterion scores in a given situation. The series of studies by Ghiselli showed that unpredictability among individuals can indeed be predicted and utilized before actual selection takes place. This means that those individuals who tend to have high absolute difference scores are classified as unpredictable and are merely left out of the prediction process. In a practical situation this obviously implies a loss of applicants who could very well turn out to be valuable employees. (Incidentally, this is also a point of critique on the off-quadrant study of Hobert & Dunnette, 1967 raised by McNemar, 1969.)

By defining unpredictability in a given situation as the algebraic difference between predictor and criterion scores, the direction of unpredictability (overpredictable or underpredictable) is determined and also can be predicted, as the results above indicate. This additional information can therefore be used effectively in prediction as discussed in the following sections.

Subgrouping to Improve Prediction

A division of the total group into three relatively independent predictable subgroups has shown to improve prediction more than the conventional regression technique previously discussed. According to Table 2 the correlations between the total on the six cognitive predictors and the criterion under cross-validation for the underpredictable, overpredictable, and accurately predictable subgroups turned out to be .26, .40, and .48, respectively. This is quite an improvement in two of the three cases on the $r_{\hat{y}y} = .34$ obtained through the regression equation involving the six cognitive predictors and the two noncognitive (biographical) scales. Even $r_{xy} = .26$ for the underpredictable subgroup is an improvement on the correlation between X and Y for the total cross-validation group which happened to be $r_{xy} = .25$. It seems, therefore, as if classifying people according to their predictive tendencies can enhance the prediction of their criterion scores to an appreciable extent. It should be noted in this regard that the classification was not based directly on the cognitive predictor composite score as in the off-quadrant study by Hobert and Dunnette (1967), but involved the development of a predictor scale against algebraic D-scores, thus evading the critique voiced by Abrahams (1969) against the mean differences between comparison groups on the predictor composite scores as employed in the Hobert and Dunnette study.

It was of some interest to know what kind of biographical items differentiated between the three groups. This information could be obtained through a content analysis of the items that predicted the three predictive tendencies well. If all the items in the predictor of error which were negatively correlated with error of prediction were endorsed by an individual, he would be classified as relatively predictable. Conversely, an individual endorsing all the items positively correlated with error of prediction would be classified as being relatively unpredictable. The direction of unpredictability (over or under) will not, however, be known from this analysis.

The items that separated the predictables from the rest vaguely describe the following kind of person: He (or she) is married, has one or more children, conservative, stable and organized, not outgoing, reserved (even subdued), not overly popular, and does not consider himself a leader.

On the other hand, the relatively unpredictable individual possesses more or less the following characteristics which are amazingly in contrast to those listed above: The person is single, has no dependents, not well organized personally, moves around frequently (with respect to residences as well as jobs), bored by unchallenging jobs, outgoing, popular, independent, considers himself a leader of others, and was considered somewhat of a problem at school.

CONCLUSION

The results obtained in this study seem to emphasize one major point, namely, that prediction of job success or academic success for that matter, can be improved beyond levels usually attainable through a more thorough study of the individuals for whom predictions are made. Interactions between the individual and the prediction instrument which make for error of prediction should not be assumed to be random. This, however, does not imply that error of measurement per se should not be considered a random phenomenon. The results of this study deal specifically with error of prediction due to systematic predictive tendencies among individuals rather than with misprediction due to random errors of measurement, which could be the result of differences in test environment and/or testing equipment and procedures. Failure to make the latter distinction is most probably directly responsible for a great deal of discord among personnel psychologists regarding the assumption of random error of measurement.

REFERENCES

ABRAHAMS, N. M. Off-quadrant comment. *Journal of Applied Psychology*, 1969, **53**, 66–68.

DUNNETTE, M. D. A modified model for test validation and selection. *Journal of Applied Psychology*, 1963, **47**, 317–321.

FISHER, J. The twisted pear and the prediction of behavior. *Journal of Consulting Psychology*, 1959, **23**, 400–405.

GHISELLI, E. E. Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology*, 1956, **40**, 374–377.

GHISELLI, E. E. The prediction of predictability. *Educational and Psychological Measurement*, 1960, **20**, 3–8. (a)

GHISELLI, E. E. Differentiation of tests in terms of the accuracy with which they predict for a given individual. *Educational and Psychological Measurement*, 1960, **20**, 675–684. (b)

GHISELLI, E. E. Moderating effects and differential reliability and validity. *Journal of Applied Psychology*, 1963, **47**, 81–86.

GHISELLI, E. E., & SANDERS, E. P. Moderating heteroscedasticity. *Educational and Psychological Measurements*, 1967, **27**, 581–590.

GLENNON, J. R., ALBRIGHT, L. E., & OWENS, W. A. *A catalog of life history items*. Greensboro, N. C.: The Richardson Foundation, 1966.

GUION, R. M. Personnel selection. *Annual Review of Psychology*, 1967, **18**, 191–216.

HOBERT, R., & DUNNETTE, M. D. Development of moderator variables to enhance the prediction of managerial effectiveness. *Journal of Applied Psychology*, 1967, **51**, 50–64.

MCNEMAR, Q. Moderation of a moderator technique. *Journal of Applied Psychology*, 1969, **53**, 69–72.

SAUNDERS, D. R. Moderator variables in prediction. *Educational and Psychological Measurement*, 1956, **16**, 209–222.

STORMS, L. H. Rationales for the "twisted pear." *Journal of Consulting Psychology*, 1960, **24**, 552–553.

TIFFIN, J., & MCCORMICK, E. J. *Industrial psychology*, Englewood Cliffs, N. J.: Prentice-Hall, 1965.

(Received December 12, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 2, 234–240

COMPARISON OF "JOINT MODERATORS" IN THREE PREDICTION TECHNIQUES

SHELDON ZEDECK,[1] C. J. CRANNY, CAROL A. VALE
AND PATRICIA CAIN SMITH
Bowling Green State University

This investigation examined the existence of potential "joint moderator" variables and compared the operation of the same potential moderator variables (anxiety and study habits) according to three different prediction techniques (subgroup analysis, differential prediction of predictability, and moderated regression). Results indicated that anxiety moderates the relationship between grade point average and American College Test scores using subgroup analysis and differential predictability techniques. A comparison of moderated regression with linear regression indicated that study habits operates as an independent predictor. None of the techniques identified joint moderators in this situation. Results are discussed in the light of problems and assumptions of each technique. The need for systematic identification of moderators from a large mass of data is indicated.

In an effort to increase predictive efficiency, investigators have given considerable attention to techniques involving moderated prediction. In particular, subgroup analysis, differential prediction of predictability, and moderated regression have been used.

The variables under investigation in these techniques are generally referred to as moderator variables. Banas (1964) defined a "moderator variable" as a quantitative or qualitative variable which improves the usefulness of a predictor by isolating subgroups of individuals for whom a predictor or set of regression weights is especially appropriate. Implicit in the moderator variable technique is the notion of dividing a heterogeneous population into homogeneous subgroups on the basis of the variables affecting the relationship between predictor and criterion. Thus, homogeneous subgroups should show differential patterns of validity.

The moderator variable concept is particularly useful in situations in which a validity coefficient is low for the entire group, but where for certain individuals, reasonably accurate prediction of criterion performance may be made from scores on the test. Applying the moderator concept allows the investigator to identify those individuals for whom more accurate predictions can be made. The practical implications of moderator techinques are increased flexibility in the use of predictors and increased efficiency of selection models, as well as the isolation of groups for whom additional or different predictors should be tried.

Prediction Techniques

Subgroup analysis. Frederiksen and Melville (1954) hypothesized that the usefulness of an interest test can be improved by identifying subgroups of engineering students for whom the test was especially appropriate as a predictor. Specifically, the usefulness of interest measures would be limited to a relatively noncompulsive group. A measure of reading speed and an unspeeded vocabulary test were used to identify compulsive and noncompulsive *Ss*. In 5 of 10 interest scales on the Strong Vocational Interest Blank, correlations between interest scores and grade point averages were significantly higher for the noncompulsive group than for the compulsive group. A replication (Frederiksen & Gilbert, 1960) yielded similar results, but only for the occupational keys most logically related to engineering.

Other bases for subgrouping have included problem-solving styles (French, 1965), job characteristics (Peterson, 1964), scores on a knowledge test (Steinemann, 1964), job satisfaction (Dawis, Weiss, Lofquist, & Betz, 1967), and anxiety and aptitude (Pervin, 1967).

Differential predictability. Ghiselli's (1956, 1960) technique involves obtaining absolute difference scores (D) between standardized criterion and standardized predictor scores. The magnitude of D serves as an index of predictability; the smaller the D, the better the relation between criterion and predictor. Correlates of the D score are subsequently identified and used as "predictors of predictability." A high positive correlation indicates that the validity should be higher for those scoring low on the "predictor of predictability" than for the entire group. The $\frac{1}{3}$ lowest and $\frac{2}{3}$ highest scorers on the predictor of predictability test are then tested for significantly different validity coefficients. A significant difference in favor of the low group indicates that the predictor is more useful for the lowest $\frac{1}{3}$ scorers than for the remaining *S*s.

---

[1] Requests for reprints should be sent to Sheldon Zedeck who is now at the Department of Psychology, University of California, Berkeley, California 94720.

Ghiselli (1956) demonstrated the potential usefulness of the technique in a study involving prediction of job proficiency of taxicab drivers. An occupational inventory differentiated those individuals for whom a tapping-and-dotting test was a good predictor of job proficiency. The measure of occupational level itself had a negligible correlation with the criterion.

Saunders' moderated multiple regression. The moderator variable is a means of maintaining the integrity of the total population while still maintaining a statistical control on each individual's membership in one of a continuous, infinite series of subpopulations defined by his score on the moderator (Saunders, 1955). Hence, Saunders (1956) treats the cross product of the moderator and predictor scores as a new predictor in any standard multiple-regression technique. The regression equation is:

$$y = \bar{y} + \sum a_i x_i + \sum b_j z_j + \sum c_{ij} x_i z_j,$$

where a, b, and c are weights, x_i the predictors, and z_j the moderators.

Saunders (1956) reanalyzed the Frederiksen and Melville (1954) data using the moderated regression technique. In 5 of 10 comparisons, moderated regression resulted in increases in predictive validity over the usual multiple-correlation method. However, in no case is there any significant improvement over the coefficients for the noncompulsive subgroup in the Frederiksen and Melville study.

Problem

The literature pertaining to moderator variables has largely ignored the effect of "joint moderators." Joint moderators may be defined as two or more variables, quantitative or qualitative, that interact to influence a validity coefficient. The traditional one-variable-at-a-time approach does not take into consideration the possibility that several moderators may have to be combined. It is conceivable that two potential moderators, for example, study habits and anxiety, will not, when operating separately, influence validity of a predictor. It is also conceivable that the validity coefficient of a predictor for a homogeneous group composed of students with high anxiety and good study habits will be significantly different from that of a homogeneous group composed of students with low anxiety and good study habits.

An approach which is relevant to the notion of moderator variables, and particularly to joint moderators, is that of configural scoring or actuarial pattern analysis (Lykken, 1956). Both moderated prediction and configural scoring identify homogeneous subgroups of individuals; moderator variables operate in the multiple-regression sense, whereas actuarial prediction permits the forming of subgroups by combining predictors in the configural sense and then investigating the criterion performance of each subgroup. This article discusses the use of joint moderators in terms of regression models.

There have been few attempts to use joint moderators. Stricker (1966), in a replication of the study by Frederiksen and Melville (1954), formed four subgroups from the combination of the two dichotomized compulsivity measures. Results indicated that the joint use of the two variables did not enhance the moderator effect. This result may be attributable to the fact that both moderators were measures of the same construct.

Klein, Rock, and Evans (1968) employed a different approach for joint moderators leading to differential prediction of law school grades. Five items (age, major field of study, degree of preparation by college for law school, father's occupation, and time at which decision was made to attend law school) were used to form multiple moderators. Essentially, individuals were placed into groups on the basis of the similarity of their profiles on a combination of items. A hierarchial grouping technique yielded four homogeneous groups. In only one of the four groups was the validity coefficient increased over the total validity coefficient.

The present investigation examined the existence of potential "joint moderator" variables in an academic setting. In addition, since the three techniques have the same intent, the operation of the same potential moderator variables was compared according to the three prediction techniques. Two variables, Study Habits Inventory (SHI; Wrenn, 1961) and Anxiety Scale Questionnaire (ASQ; Cattell & Scheier, 1963), were examined to determine their potential moderator effects on prediction of college grade point average (GPA) from the American College Test (ACT).

To determine the separate usefulness of SHI and ASQ as potential moderators, the following hypotheses were tested:

Hypothesis 1. The use of ACT scores as a predictor of GPA is differentially appropriate depending on the operation of study habits as a moderator. That is, different validity coefficients for ACT exist depending on whether *S* demonstrates good study habits (GSH) or poor study habits (PSH).

Hypothesis 2. The use of ACT scores as a predictor of GPA is differentially appropriate depending on the operation of anxiety as a moderator. That is, different validity coefficients exist depending on whether *S* is high anxious (HA), moderate anxious (MA), or low anxious (LA).

To determine the usefulness of SHI and ASQ as potential joint moderators, the following hypothesis was tested:

Hypothesis 3. The use of ACT scores as a predictor of GPA is differentially appropriate depending on the operation of a joint condition as a moderator. That is, different validity coefficients exist for each (or at least some) of the six possible joint conditions.

Method

Subjects. Test materials were mailed to 500 undergraduate students. Returns were received from 94.8% ($N = 474$) of the students who were fulfilling a research participation requirement of an introductory psychology course. Because of incomplete predictor and criterion information, the final sample size was $N = 393$.

Procedure

Subgroup analysis. To test Hypothesis 1, two study habits groups were formed using cutoffs suggested by the test manual. If *S* received $\geq +15$ on SHI, he was placed in the GSH group ($n = 209$). If *S* received $< +15$, he was placed in the PSH group ($n = 184$).

To test Hypothesis 2, three anxiety groups were formed. Less extreme cutoffs than those recommended in the manual were used so that sufficient sample sizes were obtained in the extreme groups. The HA group ($n = 104$) consisted of *S*s with scores ≥ 37; the MA group ($n = 192$) with scores ≥ 23 but ≤ 36; and the LA group ($n = 97$) with scores ≤ 22.

To test Hypothesis 3, joint moderator groups were determined based on six possible combinations of the two study habits groups and three anxiety groups resulting in the following:

Joint moderator 1 ($n = 40$): GSH and LA
Joint moderator 2 ($n = 103$): GSH and MA
Joint moderator 3 ($n = 66$): GSH and HA
Joint moderator 4 ($n = 30$): PSH and LA
Joint moderator 5 ($n = 90$): PSH and MA
Joint moderator 6 ($n = 64$): PSH and HA.

Differential predictability. All scores were converted to *z* scores. To test Hypotheses 1 and 2, D scores were correlated separately with each of the potential moderators, SHI and ASQ.

To test Hypothesis 3, D scores were correlated with each of the following two moderators: (*a*) Joint moderator A, composed of *S*s with GSH at all anxiety levels, and (*b*) Joint moderator B, based on PSH at all anxiety levels. Joint moderator score was the sum of *S*s' standardized scores on SHI and ASQ. Two independent joint moderators were used so that the compensatory effect of the two potential moderators in joint operation would be controlled and the psychological meaning of a significant effect would be facilitated.

Moderated regression. To test Hypotheses 1 and 2, SHI and ASQ, respectively, were used in separate regression equations, in a cross-product term with ACT scores. To test Hypothesis 3, two cross-product terms, ACT × ASQ and ACT × SHI, were used simultaneously in the moderated regression equation. In addition, all variables (predictor and potential moderators) were examined in a multiple regression to assess whether the variable(s) contributed to prediction in a linear sense.

Results

The validity coefficient of ACT as a predictor of GPA was .49 ($p < .01$) for the total group ($N = 393$). The standard error of estimate was .87. Table 1 presents the intercorrelations between the criterion, predictor, and potential moderator scores (single and joint).

Subgroup Analysis

Table 2 presents the correlation coefficients between ACT and GPA for the anxiety and study habits subgroups. The correlation coefficient for the MA group was significantly different ($p < .05$) from that of the HA group. The

TABLE 1

INTERCORRELATIONS

Variable	GPA	ACT	SHI	ASQ	ACT × SHI	ACT × ASQ
GPA	—	.49*	.35*	−.01	.38*	.17*
ACT			.27*	−.07	.32*	.34*
SHI				.32*	.99*	−.20*
ASQ					.32*	.91*
ACT × SHI						−.18*
ACT × ASQ						

Note.—$N = 393$. Abbreviations: GPA = grade point average, ACT = American College Test, SHI = Study Habits Inventory, ASQ = Anxiety Scale Questionnaire.
* $p < .01$.

standard errors of estimate for the two subgroups were .80 and .95, respectively. The difference between the study habits subgroups was not significant.

Table 3 presents the correlation coefficients for the joint subgroups. None of the coefficients was significantly different from any other.

Differential Predictability

Table 4 presents the correlations between the D scores and each of the potential moderators. The correlation between D and anxiety scores was significant ($p < .05$). Neither the study habits scores, nor the joint scores were correlated significantly with the D scores.

Since D correlated significantly ($p < .05$) with ASQ scores, two fractionations related to ASQ scores were made: (*a*) $\frac{1}{3}$ lowest and $\frac{2}{3}$ highest ASQ scorers, and (*b*) $\frac{2}{3}$ lowest and $\frac{1}{3}$ highest ASQ scorers. Table 5 presents the correlation between GPA and ACT for each of the fractionated groups. The difference between coefficients for the lowest $\frac{2}{3}$ and highest $\frac{1}{3}$ groups was significant ($p < .05$). The standard errors of estimate for the lowest $\frac{2}{3}$ and highest $\frac{1}{3}$ groups were .85 and .93, respectively.

Moderated Regression

Table 6 presents the multiple correlation coefficients between GPA and the predictors for both linear and moderated conditions. Item 4b indicates the relationship between GPA and ACT as affected by joint moderators; Items 2b and 3b indicate the relationship between GPA and ACT as affected by single moderators.

It is apparent from Items 2a and 4a in Table 6 that the use of SHI scores improves prediction by the linear model (increase over the simple r is significant in both cases, $p < .05$), whereas the use of ASQ scores (Item 3a) does not. Comparison of the linear and moderated models indicates no further contribution gained by the inclusion of the cross-product terms.

DISCUSSION

The essential purpose of this study was to compare the operation of potential moderators, single or joint, in three prediction techniques. The results of this study provided insights into the operation of, and problems associated with, each of the moderator variable techniques.

TABLE 2

CORRELATIONS BETWEEN GPA AND ACT FOR SUBGROUPS BASED ON ANXIETY AND STUDY HABITS

Total group ($N = 393$)	HA ($n = 104$)	MA ($n = 192$)	LA ($n = 97$)	GSH ($n = 209$)	PSH ($n = 184$)
.49	.39	.56	.45	.50	.37

Note.—Abbreviations: HA = high anxious, MA = moderate anxious, LA = low anxious, GSH = good study habits, PSH = poor study habits. Commonly underlined values are significantly different at $p < .05$.

TABLE 3

CORRELATIONS BETWEEN GPA AND ACT FOR JOINT SUBGROUPS BASED ON BOTH ANXIETY AND STUDY HABITS

Total group ($N = 393$)	PSH + LA ($n = 30$)	PSH + MA ($n = 90$)	PSH + HA ($n = 64$)	GSH + LA ($n = 66$)	GSH + MA ($n = 103$)	GSH + HA ($n = 40$)
.49	.22	.52	.30	.52	.50	.46

Note.—Abbreviations: PSH = poor study habits, GSH = good study habits, LA = low anxious, MA = moderate anxious, HA = high anxious.

Improving upon a validity coefficient that is already relatively high ($r = .49$) is both difficult and somewhat impractical. While the basic purpose of any moderator variable approach is to improve validity in situations in which predictors are poor, whether a gain is worth the effort involved in applying a moderator technique is a matter to be decided by the individual making the selection decisions.

Another problem is that of sample size, particularly in the subgroup analysis procedure. Reduction in sample size necessitated by any subgrouping approach reduces the power of a statistical test, and hence, without large initial samples it is difficult to determine the effectiveness of a moderator technique.

Anxiety operated as a "predictor of predictability" in the differential predictability technique. Improvement in validity was obtained for the $\frac{2}{3}$ lowest ASQ scorers and not for the $\frac{1}{3}$ lowest ASQ scorers. The former group included mostly LA and MA *S*s whereas the latter included mainly LA *S*s. Hence, it seems that it is the difference between HA and MA groups that was contributing to improvement. This interpretation is consistent with the results in the subgroup approach, where the significant difference between the MA and HA groups indicated that anxiety was a moderator.

Further evidence for the operation of anxiety as a moderator is found by examining the standard errors of estimate. In the subgroup analysis and the differential predictability technique, in instances where differential validities were obtained, more accurate predictions are made for the subgroup with the higher validity coefficient. However, as there is a problem of improving an already high validity coefficient, there is also the problem of improving the accuracy of prediction. In the present study, the improvement in accuracy by moderator techniques over that accuracy already existing was minimal.

Saunders' (1955, 1956) moderated regression technique provided information inconsistent with that demonstrated by the subgroup analysis and differential predictability techniques. The moderated regression technique indicated that anxiety as a predictor alone or in a cross-product term did not enhance validity. While inclusion of study habits as a predictor did enhance validity, the lack of im-

TABLE 4

CORRELATIONS BETWEEN D AND POTENTIAL MODERATOR SCORES

Group	r
D & ASQ[a]	.11*
D & SHI[a]	−.03
D & ASQ + SHI/GSH[b]	.06
D & ASQ + SHI/PSH[c]	.00

Note.—Abbreviations: D = difference scores, ASQ = Anxiety Scale Questionnaire, SHI = Study Habits Inventory, GSH = good study habits, PSH = poor study habits.

[a] $N = 393$.

[b] $N = 209$.

[c] $N = 184$.

* $p < .05$.

TABLE 5

CORRELATIONS BETWEEN GPA AND ACT FOR FRACTIONATED ASQ SCORERS

Fractionated ASQ group	r
Lowest $\frac{1}{3}$[a]	.47
Highest $\frac{2}{3}$[b]	.49
Lowest $\frac{2}{3}$[c]	.54
Highest $\frac{1}{3}$[d]	.38

Note.—Abbreviations: GPA = grade point average, ACT = American College Test, ASQ = Anxiety Scale Questionnaire. Underlined correlations are significantly different at $p < .05$.

[a] $n = 135$.

[b] $n = 258$.

[c] $n = 268$.

[d] $n = 125$.

provement in the moderated- over the multiple-regression condition indicates that study habits was acting as an ordinary predictor in the multiple-prediction sense and not as a moderator variable. Further evidence for the latter conclusion is indicated by the correlation of .99 (see Table 1) between SHI and ACT × SHI.

A possible explanation for the difference between the results for the moderated-regression technique, on the one hand, and the two approaches involving subgrouping, on the other hand, may lie in the difficulty of comparing quantitative and qualitative techniques. (Fractionating the predictability test to determine differential validities allows the treatment of the predictability variable as a qualitative variable.) The moderated-regression technique is essentially an attempt to describe the multidimensional regression surface with a model that is nonlinear in the predictors, using one equation to describe the relationship over all values of the predictor (including moderator) variables considered in the experiment. Presence of a moderator is thus indicated by the form of the equation, specifically by the significance of cross-product terms. Perhaps the moderated-regression equation should be extended to include squared and other higher order terms, thus allowing for the detectability of other forms of nonlinearity.

A subgrouping technique, however, considers each subdivision of the original population to have a different linear relationship between predictors and criteria, and moderators are detected here by differences in the correlation coefficients, which index the degree of linear relationship. The moderator is thus treated as a qualitative variable, and the effect of any nonlinearity in its relationship to the criterion and the other predictors is ignored. Close correspondence between results of the two techniques would be chiefly dependent on the felicity of choice of the composition of the subgroups.

Joint moderators were not found in this study. Thus, the a priori hypothesis that study habits and anxiety interact was not confirmed in this situation. The results do support, however, the contention that a variable can influence the relationship between a predictor and criterion without being related to either. Consequently, since discovery of moderator variables cannot be guaranteed by their relationships with other variables or on explicit theoretical bases, future research should be concerned with developing techniques for the systematic and rapid identification of moderators.

TABLE 6

LINEAR AND MODERATED CORRELATION COEFFICIENTS BETWEEN GPA AND MULTIPLE PREDICTORS

Model	R
Linear	
(1a) ACT	.49
(2a) ACT & SHI	.54
(3a) ACT & ASQ	.49
(4a) ACT & SHI & ASQ	.55
Moderated	
(1b)	
(2b) ACT & SHI & (ACT × SHI)	.55
(3b) ACT & ASQ & (ACT × ASQ)	.49
(4b) ACT & SHI & ASQ & (ACT × SHI) & ACT × ASQ	.55

Note.—N = 393. Abbreviations: ACT = American College Test, SHI = Study Habits Inventory, ASQ = Anxiety Scale Questionnaire.

REFERENCES

BANAS, P. A. An investigation of transsituational moderators. Unpublished doctoral dissertation, University of Minnesota, 1964.

CATTELL, R. B., & SCHEIER, I. H. *Handbook for the IPAT Anxiety Scale Questionnaire*. Champaign, Ill.: Institute for Personality and Ability Testing, 1963.

DAWIS, R. V., WEISS, D. J., LOFQUIST, L. H., & BETZ, E. Satisfaction as a moderator in the prediction of satisfactoriness. *Proceedings of the 75th Annual Convention of the American Psychological Association*, 1967, **2**, 269–270. (Summary)

FREDERIKSEN, N., & GILBERT, A. C. F. Replication of a study of differential predictability. *Educational and Psychological Measurement*, 1960, **20**, 759–767.

FREDERIKSEN, N., & MELVILLE, S. D. Differential predictability in the use of test scores. *Educational and Psychological Measurement*. 1954, **14**, 647–656.

FRENCH, J. W. The relationship of problem-solving styles to the factor composition of tests. *Educational and Psychological Measurement*, 1965, **25**, 9–28.

GHISELLI, E. E. Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology*, 1956, **40**, 374–377.

GHISELLI, E. E. The prediction of predictability. *Educational and Psychological Measurement*, 1960, **20**, 3–8.

KLEIN, S. P., ROCK, D. A., & EVANS, F. R. The use of multiple moderators in academic prediction. *Journal of Educational Measurement*, 1968, **5**, 151–160.

LYKKEN, D. T. A method of actuarial pattern analysis. *Psychological Bulletin*, 1956, **53**, 102–107.

PERVIN, L. A. Aptitude, anxiety, and academic performance: A moderator variable analysis. *Psychological Reports*, 1967, **20**, 215–221.

PETERSON, F. E. Identification of sub-groups for test validation research. *Journal of Industrial Psychology*, 1964, **2**, 98–101.

SAUNDERS, D. R. The "moderator variable" as a useful tool in prediction. *Proceedings, 1954 Invitational Conference on Testing Problems*. Princeton, N. J.: Educational Testing Service, 1955.

SAUNDERS, D. R. Moderator variables in prediction. *Educational and Psychological Measurement*, 1956, **16**, 209–222.

STEINEMANN, J. H. Use of a logically related predictor in determining intragroup differential predictability. *Journal of Applied Psychology*, 1964, **48**, 336–338.

STRICKER, L. J. Compulsivity as a moderator variable: A replication and extension. *Journal of Applied Psychology*, 1966, **50**, 331–335.

WRENN, C. G. *Manual, study habits inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1961. (Renewed)

(Received February 9, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 241–243

COMMUNICATING DIRECTIONAL INFORMATION WITH AN AUDITORY DISPLAY[1]

J. RICHARD SIMON[2] AND JOHN L. CRAFT

University of Iowa

This study was concerned with determining whether right or left commands could be communicated more effectively using symbolic, directional, or combined directional and symbolic cues. In a choice reaction time (RT) task, *S*s pressed a right- or left-hand key in response to: (*a*) 1,000 Hz or 4,000 Hz binaural tonal commands (symbolic cue), (*b*) the ear stimulated by a single pure tone (directional cue), and (*c*) 1,000 Hz or 4,000 Hz monaural tonal commands in the ear corresponding to the symbolic content of the command (combined directional and symbolic cues). The combined (redundant) cue condition produced the fastest RT, while RT to the directional cue alone was faster than to the symbolic cue alone.

The present study was concerned with the general problem of communicating spatial information by means of an auditory display (Gales, 1963). Specifically, this study had the twofold purpose of determining (*a*) whether simple commands (i.e., right or left) can be communicated more effectively with directional than with symbolic cues, and (*b*) whether a combination of relevant directional and symbolic cues can facilitate information processing.

There is now abundant evidence that information processing from auditory displays may be affected by irrelevant directional cues. In a study concerned with reaction time (RT) to monaural verbal commands, Simon and Rudell (1967) discovered that the speed of processing the symbolic content of a command ("right" or "left") was affected by the ear in which the command was heard. Subsequent research (Simon & Small, 1969) established that the ear stimulated provides an irrelevant directional cue which interferes with information processing on trials where it conflicts with the symbolic content of the command.

Simon (1968) attempted to facilitate the processing of verbal directional commands by using the ear stimulated to provide a relevant directional cue. He asked whether *S*s would respond faster to a display which combined directional and symbolic cues (monaural "right" and "left" commands in the ear corresponding to the content of the command) then they would to displays which involved either a symbolic cue alone (binaural "right" and "left" commands) or a directional cue alone (monaural pure tones). Results showed that, while RT to the combined cues was faster than to the symbolic cue alone, it was not faster than to the directional cue alone. Unfortunately, these data could not be used to support generalizations concerning the relative effectiveness of symbolic versus directional cues or the general effect of combining these cues. Such generalizations would have been unwarranted because qualitatively different stimuli (i.e., tones and verbal commands) were used for the different conditions. The present study obviates this difficulty by employing pure tones to present both directional and symbolic cues.

METHOD

Apparatus. The apparatus measured choice RT to a series of tones which were presented to *S* through Telephonics TDH-39, 300-ohm earphones. The earphones were mounted in NAF-48490-1 cushions and fixed to a standard headband. The *S*s' task was to press the correct one of two finger keys as quickly as possible after hearing the tone. Two Hewlett-Packard Model 200AB audio oscillators were used to generate 95-db. tones of 1,000 and 4,000 Hz. Silent switches permitted *E* to present either tone to the left or right ear alone, or to both ears simultaneously. A Hunter Klockounter started when the tone was presented and stopped when *S* pressed a key. Depressing the key also turned off the tone and signaled *E* as to which key (or keys) had been pressed. The keys, which *S* operated with his right and left index fingers, were mounted 10 in. apart on a table in front of him.

Subjects. The *S*s were 36 male and 36 female University of Iowa undergraduate volunteers. All *S*s were strongly right-handed and reported normal hearing.

---

[1] This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to J. Richard Simon, Department of Psychology, University of Iowa, Iowa City, Iowa 52240.

TABLE 1

Reaction Time (in Msec.) to Tonal Directional Commands as a Function of Cue Properties of the Stimulus and Responding Body Member

Cue	Responding hand		
	Left	Right	$\bar{X}$
Symbolic (S) only	368	374	371
Directional (D) only	309	298	304
D & S combined	297	293	295
$\bar{X}$	325	322	

Procedure and experimental design. Each *S* performed on three blocks of trials corresponding to the three experimental conditions. A 3 × 3 Latin square was used to balance the order of presentation of the conditions. Twelve males and 12 females were randomly assigned to each of the three sequences.

In the block of trials involving the symbolic cue, *S*s were told that they would hear either a high-pitched tone or a low-pitched tone in both ears simultaneously. Half of the *S*s were instructed to press the right key if they heard the high-pitched tone and the left key if they heard the low-pitched tone. The other half of the *S*s were given the opposite pitch-key association.

In the block involving the combined directional and symbolic cues, half of the *S*s were told that they would hear either a high-pitched tone in their right ear or a low-pitched tone in their left ear. Instructions stressed the fact that the high tone would always appear in the right ear, and the low tone would always appear in the left ear. They were told to press the right key if they heard the high-pitched tone in their right ear and to press the left key if they heard the low-pitched tone in their left ear. The other half of the *S*s performed using the opposite pitch-key association. Each *S* used the same pitch-key association on this block as he did on the block involving the symbolic cue.

In the block of trials involving the directional cue, *S*s were told that they would hear a tone in either their right or left ear and that they should respond by pressing the corresponding (right or left) key. For half of the *S*s, the 1,000-Hz tone was used for the first half of the test trials and the 4,000-Hz tone was used for the second half. The remaining *S*s heard the 4,000-Hz tone during the first half of the block and the 1,000-Hz tone during the second half. The *S*s were simply told half way through the trial block "I am now going to change the pitch of the tone. Your task, however, is the same as before." This procedure was followed in order to control for possible effects of tone frequency per se on RT. Thus, the directional, symbolic, and combined directional and symbolic blocks were all comparable in terms of the number of high- and low-pitched tones presented.

Each block of trials consisted of a predetermined random sequence of 44 tonal commands, half of which called for a right response and half a left response. The first four responses in each block, two right and two left responses, were practice trials. A warning light was presented 1 sec. prior to the onset of each trial, and there was a 9-sec. interval between trials.

Results

Median RTs were computed for each *S* for both his right- and left-hand responses under the three treatment conditions (i.e., the symbolic, the directional, and the combined directional and symbolic blocks). Table 1 summarizes the means of these median RTs.

An analysis of variance revealed a highly significant overall main effect of cue ($F = 226.02$, $df = 2$, 132, $p < .001$). It can be seen in Table 1 that RT to the symbolic cue (371 msec.) was markedly slower than to the directional cue (304 msec.) and to the combined directional and symbolic cues (295 msec.). However, the analysis of variance also revealed a significant Cue × Hand interaction ($F = 5.24$, $df = 2/132$, $p < .01$). Table 1 shows that while RT was faster for the right- than the left-hand response under the directional and combined cue conditions, the reverse was true under the symbolic cue condition. Additional analyses demonstrated that this reversal of the hand effect under the symbolic cue condition was the major cause of the overall Cue × Hand interaction; that is, when only the directional and combined cue conditions were included in the analysis, the Cue × Hand interaction was no longer significant, ($F = 1.73$, $df = 1/132$). Furthermore, the fastest RT to the symbolic cue (368 msec.) was still significantly slower than the slowest RT (309 msec.) to either of the other two cue conditions ($F = 114.76$, $df = 1/132$, $p < .001$). Thus, on the basis of these findings, it was decided that the overall Cue × Hand interaction did not preclude looking at the simple main effects of cue.

Mean comparisons revealed that RT to the directional cue was faster than to the symbolic cue ($F = 295.98$, $df = 1/132$, $p < .001$), and RT to the combined directional and symbolic cues was faster than RT to the directional cue ($F = 5.34$, $df = 1/132$, $p < .025$).

The analysis revealed two additional significant sources of variance. Males responded faster than females ($F = 7.65$, $df = 1/66$, $p < .01$), and RT decreased over successive blocks of trials ($F = 12.73$, $df = 2/132$, $p < .001$).

DISCUSSION

Two results of the present study are worthy of note. One is the finding that directional cues were far more effective than symbolic cues for presenting simple right and left commands. Information was decoded much more rapidly when the cue was provided by the ear stimulated than when it was provided symbolically (i.e., as a frequency discrimination). This finding underlines the potential practical value of utilizing directional cues to provide spatial information in auditory displays. It also emphasizes again that the concept of S-R compatibility (Fitts & Seeger, 1953) is as applicable to auditory displays as to visual displays (Simon, 1967).

A second finding is that information processing was, statistically, significantly faster when relevant directional and symbolic cues were both present in the display than when either cue alone was present. This finding fits into a small body of literature which suggests that, under certain conditions, providing redundant information in a display or simultaneous signals to different modalities may facilitate information processing (McCormick, 1964; Woodworth & Schlosberg, 1954). Because the observed difference was so small, however, this finding is of perhaps greater theoretical than practical significance.

REFERENCES

FITTS, P. M., & SEEGER, C. M. S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 1953, **46**, 199–210.

GALES, R. S. Auditory presentation of information. In C. T. Morgan, J. S. Cook, A. Chapanis, & M. W. Lund (Eds.), *Human engineering guide to equipment design*. New York: McGraw-Hill, 1963.

MCCORMICK, E. J. *Human factors engineering*. New York: McGraw-Hill, 1964.

SIMON, J. R. Choice reaction time as a function of auditory S-R correspondence, age and sex. *Ergonomics*, 1967, **10**, 659–664.

SIMON, J. R. Reaction time as a function of the cue properties of an auditory display. *Journal of Applied Psychology*, 1968, **52**, 224–226.

SIMON, J. R., & RUDELL, A. P. Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 1967, **51**, 300–304.

SIMON, J. R., & SMALL, A. M., JR. Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, 1969, **53**, 433–435.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.

(Received December 4, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 244–250

LINGUISTIC DESCRIPTION OF AUDITORY SIGNALS[1]

PATRICIA WRIGHT[2]

Medical Research Council, Applied Psychology Unit, Cambridge, England

In three experiments, *S*s gave verbal descriptions of tones differing in pitch, intermittency, and modulation, and the extent of the agreement between *S*s was noted. In Exp. I there was greater consistency across *S*s when descriptions referred to the tone's physical characteristics than when onomatopoeic and illustrative terms (e.g., bleep, hooter) were used. Onomatopoeic labels appear to convey no more specific information than general terms (e.g., sound). However, even when labeling physical characteristics there was usually high agreement for only one feature of the stimulus, although the salient dimension varied with different signals. A multiple-choice procedure in Exp. II showed that if *S*s had to label each tone along two physical dimensions, the consistency between *S*s increased in accordance with the predictions of a model which assumes that judgments are made along each dimension independently, then combined probabalistically. Experiment III more closely examined the distribution of verbal labels along the pitch dimension. The data indicated 95% agreement between *S*s for the description of isolated pure tones (heard only once and therefore without background or reference tones) if these were above 600 Hz or below 200 Hz.

Although the meanings of many auditory signals are easily learned, there are some occasions when people must be told before they ever hear the signal that sound X means embark upon a particular course of action. Fire alarm signals are a typical example. International telephoning is another (Shoeffler & Sherridan, 1966). In this latter case the caller is faced with a completely new range of auditory stimuli among which he must discriminate.

Perhaps the simplest way in which telephone users can be provided with the necessary information is for verbal descriptions of the different tones to be included in the general issue of telephone directories or handbooks, or displayed alongside telephones. But the problem arises of how such signals can be described in order that the correct identification of the sound is a near certainty. The following experiments examine various aspects of the verbal labeling of telephone-like sounds. The first experiment is concerned with how people spontaneously label these sounds.

EXPERIMENT I

Procedure. Answer papers were distributed to *S*s and the following instructions were given:

> This experiment is concerned with how people describe sounds. I want you to imagine that you have to leave a written message for someone, a message telling them to listen for a certain sound then do something. For example, "When you hear the bell it's tea time" or "When you hear the hooter pack up and go home," etc. This message is written at the top of your paper, only the words saying what you have to listen for are missing. (The message written was "When you hear . . . do something.") Your task will be to supply these words as you listen to sounds on a tape recorder. You can use as few or as many words as you like, and each sound will be repeated several times, but the gap between different sounds is rather short so don't wait for a sound to finish before you start writing. Start writing while you can hear the sound. Any questions?

Because it was clear that some *S*s had opted out of the experimental situation and were not writing anything at all, the tape was stopped after the nineteenth tone, and *S*s were instructed to "think of the tones as being call tones in a hospital. Imagine that you're writing a roster of which call refers to which doctor and try to describe the tones with this in mind." This reorientation had the desired effect of increasing the number of written descriptions.

The data were obtained from a group of 17 adult *S*s who were paid for taking part in the experiment.

Materials. There were 31 tape-recorded signal tones. For each signal, the burst of tone lasted 20 sec., with 5 sec. between different tones. However, because of the difficulty in persuading *S*s to describe the tones, analysis has been confined to the last 12 signals. The physical description of these signal tones is given in Table 1.

---

[1] This article has been accepted for publication by the outgoing Editor, Kenneth E. Clark.

[2] Requests for reprints should be sent to the author, Medical Research Council, Applied Psychology Unit, 15 Chaucer Road, Cambridge, England.

TABLE 1

PHYSICAL CHARACTERISTICS OF THE SIGNAL TONES

Signal tone no.	Frequency (in Hz)	Amplitude modulation (in Hz)	Time on (in sec.)	Time off (in sec.)
20	310	0	1.00	1.00
21	850	32	1.50	1.50
22	425	16	.15	.15
23	620	16	.33	.67
24	425	0	2.00	4.00
25	620	50	.50	.50
26	620	16	.03	.27
27	620	50	.60	5.40
28	425	50	.03	.27
29	425	16	.60	5.40
30	425	0	.50	.50
31	440 + 480	0	2.00	4.00

Results

It was immediately apparent from the data that *S*s had no readily available vocabulary to apply to auditory signals such as telephone tones. Many *S*s used onomatopoeic terms such as bleep, buzz, pop, umph, etc. Others made use of illustrations, for example, horn type noise, whistle, siren, etc. These two categories were usually, but not necessarily, mutually exclusive, for example, "The buzz of an electric door bell." Nor was it necessary to use either category, for example, short high-pitched tone. But of 204 opportunities for response, terms that were largely onomatopoeic were used 104 times, there being 18 different terms within this category; illustrative examples were given 101 times, there being 36 different illustrations; and the generic terms such as tone, note, or sound—or even no term at all—occurred 34 times. Clearly, the onomatopoeic and illustrative categories predominated. However, there was little consistency across *S*s in their use of these labels. Table 2 shows the frequency distribution of the five most common responses in both the onomatopoeic and the illustrative categories.

Table 2 shows that response terms are used with very little discrimination. For example, each of the 12 signal tones was at some time described as a bleep, and even within each of the two categories there were usually three or more labels being used to describe a single tone.

Sometimes *S*s referred to the physical characteristics of the sounds, such as high or low, long or short, fast or slow. Table 3 shows the frequency distribution of references to these characteristics.

TABLE 2

FREQUENCY OF THE FIVE MOST COMMON DESCRIPTIONS USING ONOMATOPOEIA AND ILLUSTRATIONS

Signal tone no.	Onomatopoeic descriptions					Illustrative descriptions				
	Bleep	Buzz	Pip	Brr	Blip	Morse code	Hooter	Engaged tone	Ringing tone	Siren
20	2	3	2	—	—	—	2	2	—	1
21	5	2	1	—	—	—	—	—	2	—
22	4	—	3	3	—	2	—	—	1	—
23	3	—	1	1	2	—	—	—	1	—
24	4	1	—	—	—	—	1	1	—	1
25	3	—	—	—	1	1	1	2	—	2
26	3	—	2	—	1	3	—	—	—	—
27	3	3	—	—	—	—	2	—	—	2
28	1	—	1	—	—	3	—	—	—	—
29	2	3	—	3	—	—	1	1	4	—
30	2	2	—	—	1	2	1	2	—	—
31	1	3	1	3	—	—	3	1	1	—
Total	33	17	11	10	5	11	11	9	9	6

TABLE 3

FREQUENCY OF REFERENCE TO PHYSICAL CHARACTERISTICS OF THE SIGNAL TONES

Signal tone no.	Pitch			Rate			Duration			Interval		
	High	Medium	Low	Fast	Medium	Slow	Short	Medium	Long	Short	Medium	Long
20	—	1	11[a]	—	—	3	—	3	2	1	2	1
21	14[a]	—	—	—	—	—	—	—	6	3	—	1
22	—	2	2	10[a]	—	—	3	—	—	—	—	—
23	9[a]	1	—	4	—	—	4	—	—	3	1	—
24	—	2	9[a]	—	—	2	—	—	11[a]	—	—	3
25	4	2	—	4	1	—	2	—	—	—	—	1
26	1	1	1	10[a]	—	—	2	—	—	—	—	—
27	1	3	1	—	—	4	4	—	3	—	—	11[a]
28	—	—	9[a]	7	—	—	—	—	4	—	—	—
29	—	1	1	1	—	6	3	—	1	—	—	7
30	4	1	—	5	1	—	1	1	3	2	1	—
31	2	2	2	—	1	1	—	2	8	—	1	2
Total	35	16	36	41	3	16	19	6	38	9	5	26
No. references to physical characteristic		87			60			63			40	

[a] More than half the *S*s gave this response.

From Table 3 it can be seen that for most signal tones there was one predominant response; for example, Tone 21 was high, whereas Tone 22 was fast. The notable exceptions were Tone 24—which was both long and low—and Tones 29, 30, and 31—for which no single response was given by more than half the *S*s.

It might be expected that if verbal labels such as bleeps or pips had been used, this lessened the need to use descriptions such as high and short. For example, a series of high, short tones might be synonymous to a series of bleeps. This possibility was examined by comparing the frequency with which various physical characteristics (pitch, rate, duration, interval) were mentioned when *S*s used the label "tone" and when they used the label "bleep" (this being the most common of the onomatopoeic labels). The average number of characteristics mentioned in conjunction with bleep was 1.73 and in conjunction with more general terms was 1.60. Thus, the present data do not support the suggestion that onomatopoeic terms provide a shorthand encoding a sound's physical characteristic.

Sometimes no nouns were used in labeling a signal, instead people just wrote high pitched, fast. On such occasions the average number of characteristics mentioned was 2.58, but for the present data it is not clear whether there was a general tendency to describe the signal tone more fully when not categorizing it as a noun of some kind, or whether the difference was primarily a function of individual variation in labeling techniques, for a few people tended to omit nouns in most of their responses, others used illustrations for the majority of signals.

On only a few occasions was there any attempt to apply a verbal reference to the modulation of the signal tone, and then terms such as "wavery" were used.

Discussion

From the difficulty experienced in persuading *S*s to write down descriptions of the tones, it was evident that they did not have verbal labels readily available for the kinds of signal tones used in this experiment. The results bear this out. Consistency across *S*s tended to be low, particularly when people used either illustrative lables or onomatopoeic words. However, the latter did not appear to convey any specific information since the number of adjectives used with these terms was no less than the number used with more general terms such as "tone."

This lack of agreement about onomatopoeic descriptions might seem unexpected in view of the likelihood that phonetic symbolism should be relatively strong in this field. Weiss (1964) quotes Jesperson to the effect that words related to sound would show more phonetic symbolism than many other words. However, it is possible that the effects of phonetic symbolism are manifest only under a specified range of experimental conditions (Weiss, 1964), and the present situation might be too insensitive to detect such effects.

That consistency between *S*s in the use of illustrative examples is also low and supports the findings of Corcoran, Carpenter, Webster, and Woodhead (1968). Corcoran et al. showed that providing listeners with illustrative descriptions leads to poorer identification of engine noises than does the provision of information about the physical characteristics of the sounds. Clearly, what is "morse dots" to one man may be "a series of pips" to another, and the terms do not serve as a useful basis for interperson communication.

Although agreement between *S*s was highest when physical attributes were mentioned, there remain the problems of which physical dimensions it is possible to code verbally without losing this consistency, and also how many such dimensions of a particular stimulus can usefully be incorporated into a single description (i.e., Is "a high, loud, fast, wavery tone" comprehensible?).

On the first of these problems, data are available from a study by Markel (1965) which suggest that there is high agreement between *S*s along the dimensions of pitch, loudness, and tempo. The data from the present study substantiate this, since pitch and tempo, or rate, were the main physical categories referred to.

On the second problem raised above, the present data suggest that most people seem to take cognizance of only one physical dimension of the sound. If it should be the case that the majority of untrained *S*s can only refer a complex signal to one dimension, than this severely limits the number of different tones among which it is feasible to ask people to discriminate. It therefore seemed necessary to examine whether or not *S*s could be persuaded to use more than one physical dimension for categorizing the auditory stimuli.

Pitch	Rate	
High	Fast	
	Medium	
	Slow	
Medium	Fast	
	Medium	
	Slow	
Slow	Fast	
	Medium	
	Slow	
None of the above		
Cannot tell		

FIG. 1. Illustration of page from answer booklet used in Exp. II.

EXPERIMENT II

Procedure and materials. Another 22 adult *S*s were presented with the same signal tones used in Exp. I, but this time they were given answer booklets (see Figure 1). In half the books the heading Duration was replaced by Rate and the rows were labeled fast, medium, and slow.

The *S*s were instructed to write the number of each tone on the row they thought that it belonged, and were told that there were no restrictions on how often the same row could be used. The numbers 1–31 were written on a blackboard as the tones were presented, one number for each tone, and *S*s used these numbers to designate the tone to which they were listening. After every fifth tone, *S* turned the page of the answer booklet. This was done to overcome any reluctance there might be in using a row that already contained a number.

Results

The consistency between *S*s is shown by Table 4 which gives the percentage of *S*s who gave the modal response on each dimension; that is, 100% indicates perfect consistency.

TABLE 4

Percentage of *S*s Giving the Modal Response in the Multiple-Choice Task

Signal tone no.	Pitch-Rate			Pitch-Duration		
	Pitch	Rate	Both	Pitch	Duration	Both
20	82	73	64	91	55	45
21	82	55	55	91	73	73
22	82	100	82	82	100	82
23	100	64	64	100	64	64
24	64	100	64	82	91	73
25	73	55	36	82	55	45
26	73	91	64	45	91	45
27	73	91	64	64	45	27
28	73	100	73	91	100	91
29	73	64	45	82	55	36
30	45	55	27	82	55	55
31	45	100	45	45	100	45
$\bar{X}$	72.1	78.0	56.9	78.1	73.7	56.8

Table 4 shows that there is very little difference between the combination pitch-rate and pitch-duration. It would also appear from the column means that *S*s can use two dimensions to label auditory stimuli. This latter finding is a little misleading, for although 75% of the *S*s could agree that a tone was, say, fast and 75% of them would agree that it was high, only 50% would have agreed to the tone being called high and fast. Nevertheless, there was rarely anyone who did not think that either it was high or it was fast. This is illustrated diagrammatically in Table 5.

Consequently, there is a sense in which labeling the stimulus tone along two dimensions lessens the divergence among *S*s. This is shown by comparing the number of people who gave a response outside the modal category in Exp. I and II (see Table 6). More people gave a nonmodal response in Exp. I than in Exp. II ($p < .01$, Fisher Yates Exact Test). This was true both for pitch-rate and pitch-duration, which suggests that although *S*s spontaneously labeled only one physical dimension, many *S*s were able to label two such dimensions.

TABLE 5

Typical Percentage of Distribution of Correct Responses to a High, Fast Signal Tone

Rate	Pitch		Total
	High	Not high	
Fast	50	25	75
Not fast	25	—	25
Total	75	25	

TABLE 6

Percentage of *S*s Not Giving the Modal Response

Signal tone no.	Exp. I	Exp. II[a]	
		Pitch-Rate	Pitch-Duration
20	35	0	0
21	18	9	0
22	41	0	0
23	47	0	0
24	35	0	9
25	76	9	9
26	41	0	0
27	35	0	18
28	47	0	0
29	59	9	0
30	71	27	18
31	53	18	0
$\bar{X}$	46.5	4.7	3.5

[a] *S*s who gave a response that did not include the modal response on either dimension.

Discussion

The data show that drawing people's attention to the physical characteristics of the stimulus and specifying the features to which they must attend, increases the consistency between *S*s, even when considering the responses along a single dimension. Averaging the items with superscripts in Table 3 indicates approximately 60% agreement between *S*s in Exp. I; Table 4 shows that in Exp. II agreement was over 70% within each dimension. Probably this is the result of some *S*s being able to label features such as pitch and rate, although they do not always do so spontaneously. Nevertheless, the degree of consistency is considerably less, dropping to 50%, when the modal response is defined with reference to two dimensions. Therefore, referring to two dimensions of a signal tone will not inevitably result in more accurate identification of the signal.

However, if the stimuli to be discriminated happen to differ along two dimensions, then labeling both dimensions can be expected to reduce the number of people who select the incorrect signal (e.g., Table 6 shows that this reduction was from over 40% incorrect choices, obtained when using pitch alone, to less than 5% incorrect when using pitch together with either rate or duration). Thus, reference to two dimensions is more useful for the response possibilities which it excludes rather than for those which it includes.

The present data on the effect of combining two dimensions can be accounted for in terms of the model for recognition of two-dimensional auditory stimuli proposed by Corcoran (1967). This model, termed the independence model, predicts that if an observer's judgment of one stimulus parameter (e.g., pitch) is uninfluenced by the value of the other parameter (e.g., rate), then the probability of perceiving the whole stimulus correctly is simply the product of perceiving each part correctly. From Table 4, it is clear that the both columns approximate the products of the other two columns.

EXPERIMENT III

One factor which might limit the generality of the findings from the two previous experiments is that in both studies *S*s heard a range of signal tones. It is possible that the first few tones might create a frame of reference against which later signal tones are evaluated. This in turn might mean that the observed agreement between *S*s is spuriously high, being due largely to all *S*s using the same terms of reference.

In order to check this, and also to obtain more detailed information on the distribution of linguistic categories within the pitch dimension, *S*s were asked to label the pitch of pure pure tones heard only once.

Procedure and materials. Eight groups of adult *S*s (numbering 13–17 *S*s per group) were told to imagine that they had just dialed a telephone number, that they were listening for a signal tone, and that they then heard the sound that was about to be played to them. The *S*s were instructed to write down whether this sound was high or low. A tape which recorded 22 sec. continuous pure tone was then played. The pitch of this tone was different for each of the eight groups. No *S* heard more than one tone.

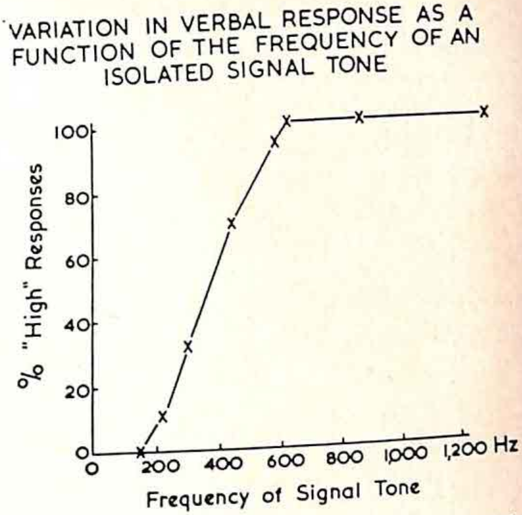


FIG. 2. Variation in verbal response as a function of the frequency of an isolated signal tone.

Results

The percentage of *S*s in each group who responded "high" is shown in Fig. 2 for each of the signal tones.

Clearly, there is agreement between *S*s that tones above 600 Hz are high, and those below 200 Hz are low.

Discussion

The present results show clearly that agreement between *S*s was not dependent upon a frame of reference having been introduced into the experimental situation. People would appear to have established reference points, at least along the pitch dimensions, prior to entering the laboratory. Moreover, there was fairly wide agreement on what these reference points are (viz, 200 and 600 Hz).

All this is not to argue that the label given to a signal is invariant over contexts, there is ample evidence to the contrary (cf. the effects of context, duration, intensity; Woodworth & Schlosberg 1954, pp. 229, 335.), but rather to suggest that when there is no readily available context, then a tone above 600 Hz can confidently be described as high.

Perhaps there are also reference points common to the majority of adult *S*s in other physical dimensions (e.g., rate, modulation, etc.). Little seems to be known about this aspect of verbal communication, about what stereotypes are common to adults, and how widely any particular commonality extends particularly across occupations. (The majority of *S*s in the present experiments were housewives.)

In conclusion, the results of the present studies suggest that, in general, *S*s do not have a readily available vocabulary for describing auditory signals such as those used by telephone companies, although there may be population stereotypes for some physical characteristics of such tones, and it is possible to direct people's attention to the relevant physical dimensions.

REFERENCES

CORCORAN, D. W. J. Perceptual independence and recognition of two-dimensional auditory stimuli. *Journal of the Acoustical Society of America*, 1967, **42**, 139–142.

CORCORAN, D. W. J., CARPENTER, A., WEBSTER, J. C., & WOODHEAD M. M. Comparisons of training techniques for complex sound identification. *Journal of the Acoustical Society of America*, 1968, **44**, 157–167.

MARKEL, N. The reliability of coding paralanguage. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 306–308.

SCHOEFFLER, M. S., & SHERIDAN, J. A. Laboratory studies related to subscriber international dialing. Part 2: Confusability of audible tones. Murray Hill, N. J.: Bell Telephone Laboratories, 1966. (Mimeo)

WEISS, J. H. Phonetic symbolism reexamined. *Psychol logical Bulletin*, 1964, **61**, 454–459.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimenta psychology*. Methuen, 1954.

(Received November 14, 1969)

Journal of Applied Psychology
1971, Vol. 55, No. 3, 251–258

BINOCULAR COORDINATION IN READING[1]

KARL U. SMITH,[2] ROBERT SCHREMSER, AND VERNON PUTZ

Behavioral Cybernetics Laboratory, University of Wisconsin

This experiment measured the differences in time between the two eyes, using laboratory real-time computer methods to detect and measure the time between velocity peaks of binocular saccadic movements in reading. The hypothesis was that instead of being completely conjugate as indicated by prior methods of ocular measurement, the eyes must be coordinated in directional motion by small time differences that govern their feedback guidance and relative velocity. Results indicated that the time differences between the eyes clustered around three values: (*a*) near synchrony, including no difference and left eye leading by 1 msec.; (*b*) left eye leading by 7–9 msec.; and (*c*) left eye leading by 14 msec. These time differences were not related to the difficulty of the reading material, but were changed significantly by 15° horizontal rotation of the reading display. The results change the established views that the eyes are perfectly conjugate in saccadic motion and provide initial suggestive data toward a dynamic feedback doctrine of coordinate eye motion and functional disabilities in visual perception.

The classical view defined by von Helmholtz (1924–1925) is that the two eyes are exactly conjugate in space and form perception. The modern statement of this view holds that exact time synchronization is characteristic of both saccadic and pursuit movements (Alpern, 1962). The only dissent from this classical belief has been from Stromberg (1938), who found large differences between the positions of the two eyes prior to and after saccades. Alpern (1962) and Yarbus (1967) reported near perfect identity of timing of the two eyes in binocular motion. Yarbus, however, confirmed Stromberg's observations on differences in eye positions, but concluded that these differences were related to the fact that a change of fixation in binocular space perception is composed of two independent movements—that is, convergence and divergence and a saccade—but that binocular movements during a saccade are identical.

The problem of synchronism of the two eyes in different types of movement—saccadic, pursuit, divergence, and convergence—represents an almost unexplored field because exact methods for recording binocular eye movements and timing them have not been developed. However, in vision, as in all other psychophysiological systems, exact timing may be assumed to be fundamental to all critical dynamic operations related to the control of velocity and integration of movements. The view behind this research is that feedback timing—the interval defining the dynamic closed-loop control of retinal input by ocular motions—is just as critical as ocular and optometric factors in governing space perception. This hypothesis was suggested initially by results of experiments which showed that eye-movement-retinal feedback delays of .1 sec. and less impaired vision and accuracy of eye tracking (Smith & Molitor, 1969; Smith, Putz, & Molitor, 1966).

The present experiment has been conducted to develop real-time laboratory computer methods to measure the time differences between the two eyes with high precision, and thus to evaluate the classical view that there is exact synchronism between binocular saccadic movements. The creation of an analog-digital-analog computer system as a machine for dynamic vision research has made possible many detailed types of experimental measurements on vision, of which measurement of binocular time synchronism is only one of many possible innovations (Smith, 1966; Smith & Henry, 1967; Smith, Putz & Molitor, 1962).

---

[1] This research was supported by grants from the National Science Foundation and the Social and Rehabilitation Administration and was carried out in the context of a training program on psychophysiological cybernetics sponsored by the Biological Training Section of the National Institute of Mental Health.

This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to Karl U. Smith, Behavioral Cybernetics Laboratory, University of Wisconsin, 918 Conklin Court, Madison, Wisconsin 53715.

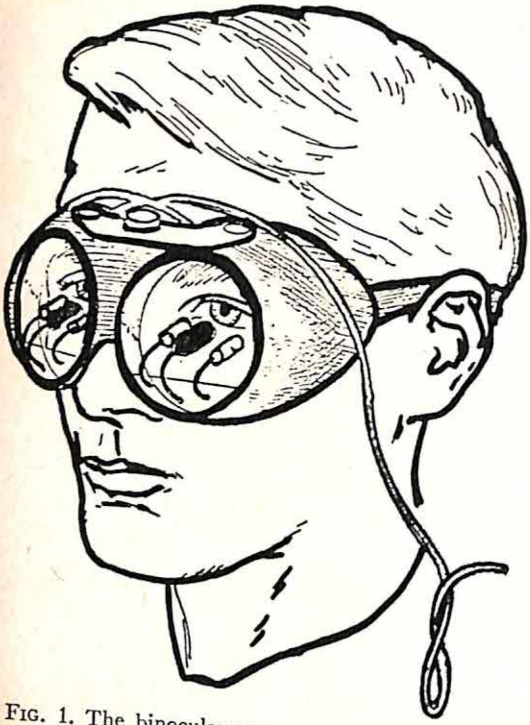


FIG. 1. The binocular eye-movement transducer.

METHOD

Some of the temporal features of saccadic movements are common knowledge. The saccadic movement is the rapid ballistic movement of the eye which typically occurs between fixations. Once a saccade is initiated, its amplitude cannot be altered by a second stimulus for fixation. The velocity change in eye motion during a 15°–20° saccade is approximately sinusoidal—i.e., the increase and decrease in velocity during the course of the movement are approximately equal. In reading, very small saccades of .5°–5° occur between fixations within a single line of print, with somewhat longer movements in shifting from the end of one line back to the beginning of another. The latency of saccadic movements varies considerably, depending on whether they occur singly or successively in a line. Roughly, a first saccade may have a latency of around 200 msec., with latencies of subsequent in-line movements being reduced to less than a third of that value. The duration of the movement of the saccade may vary considerably, from a few to over 100 msecs.

The specific object of this study was to measure the time differences in the saccadic responses of the two eyes in reading. The task of *S*s was to read three passages of varying difficulty under three conditions in which the reading display was oriented normally, rotated horizontally 15° to the right, and rotated horizontally 15° to the left. While reading, *S* wore a pair of binocular photoresistor eye-movement recorders which transduced the separate eye movements to electrical analog signals that were recorded on magnetic tape. At the end of a series of experimental sessions, the data from these tapes were processed by an analog-digital-analog computer system and a special experimental program used to detect the occurrence of paired binocular saccadic movements and to measure the time differences between them in milliseconds. The critical innovations in method in this experiment were in using the velocity signal of eye movements to establish their real-time occurrence with precision and the use of experimental computer programming and high rates of digital sampling to measure the interval between the paired binocular saccades.

The binocular eye-movement transducers used in the study are shown in Figure 1. These transducers are composed of a light source and two photoresistors for each goggle aperture. The light and the photoresistors are located below the line of sight within each aperture so that *S* has clear vision in each eye. The small red lights illuminate the corneas. The photoresistors, which are arranged in a bridge circuit, sense variations in the relative amount of light reflected from the cornea as the eye moves in a horizontal plane.

The experiment involved two phases of computer control of experimental operations. In the first phase, the hybrid computer system was used to schedule trials and to control recording of the eye-movement data on magnetic tape. The arrangement of the components of the computerized vision laboratory are shown in Figure 2. The *S* was seated comfortably in a modified dental chair and arranged so that the head would not be moved after the eye-movement transducers were arranged in place. The reading material was located directly in front of *S*s eyes at a distance of 24 in. An oscillograph was used to monitor the experiment and to control the recording of the eye movements on magnetic tape. The 160-A (Control Data Corporation) digital computer was programmed to generate trial marks which were recorded both on the oscillograph and on the magnetic tape recorder. The recorded marks on the oscillograph were used by *E* to control presentation of trials in reading, while those on the magnetic tape were used to identify trials in the recapturing of data.

As shown in Figure 2, the electrical signals from each eye-movement transducer were regulated by an electrical bridge and gain control and then amplified. The amplified signals were then recorded on two channels of the oscillograph and on separate channels of the magnetic tape recorder. These same signals also were transmitted to differentiators and these differentiated velocity signals recorded on the oscillograph. The digital computer was programmed to sequence the order of trials and to generate signals to code these trials on a third channel of the magnetic tape recorder.

After a series of trials for a particular *S* was run, the data for these trials were processed by the computer system to measure the time differences between paired saccades of the binocular eye movements. These measurements were made by determining the time interval between the peak velocity of paired saccades. Figure 3 indicates how the components of the computer system of the vision laboratory were arranged to make these measurements. The analog signals of the paired eye movements, as regenerated by the tape recorder, were amplified and differentiated. These differentiated velocity signals were then multiplexed for digital conversion. An analog-digital computer converted the

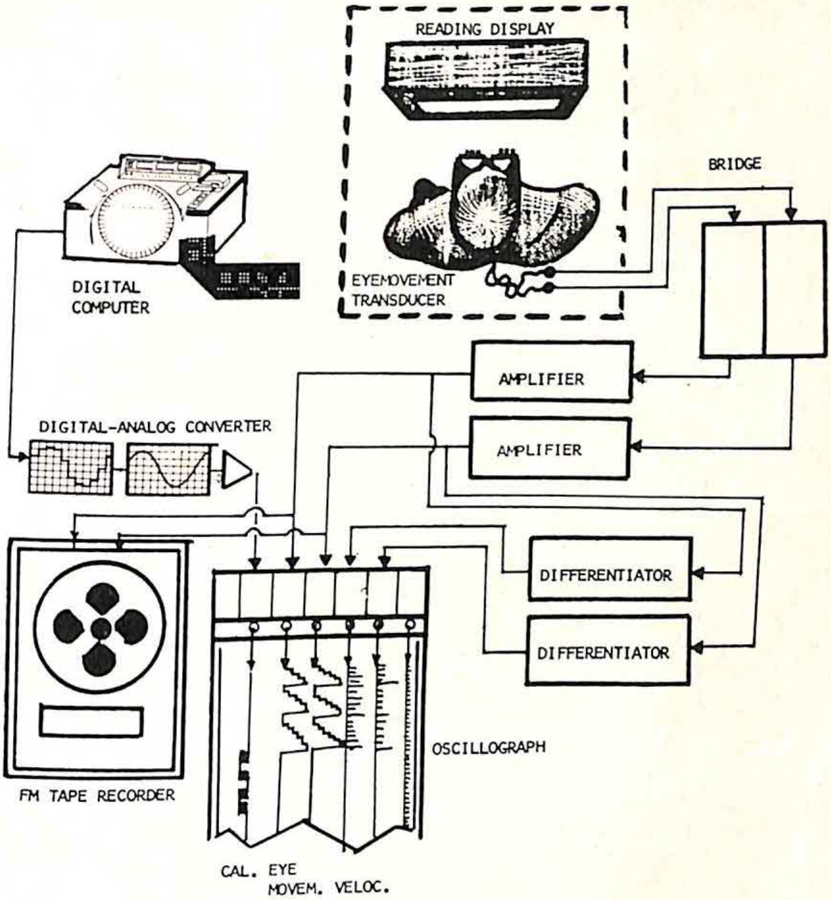


FIG. 2. Components of the real-time computer system used in recording eye movements.

multiplexed eye-movement signals to digital form at a sampling rate of 1,000 samples per second and measured each of these signals on a nine-bit digital scale. These measured signals were then selected by the digital computer for the differential time measurement.

The digital computer was programmed to detect the peak velocity value of the two digitized eye-movement velocity signals and to compute the time interval between the two values at the sampling rate used (i.e., 1,000 samples per second). The computer program also included procedures for indexing these computed time measures and for storing them in memory. After the magnetic tape record for a complete series of trials was run, the time data were punched out in digital form on paper tape and in a form for later statistical analysis.

The actual program for the time measurements operated in terms of several steps. The initial velocity blip of the two digitized movement signals was assigned a value of 0000 and the computer thereafter searched for the occurrence of a second peak value in the incoming signal. If this second peak value occurred within a predetermined elapsed time of 20 msec., the sampled value of this elapsed time was indexed and stored. For example, if the second peak velocity value occurred 5 msec. after the first peak, a value of 0005 was assigned this second peak value. However, if the second peak value occurred in the same signal channel as the first peak, the program blocked and stopped scoring a particular trial. Eye blinks typically caused such outlying data and accounted for most of the failures to obtain data for a number of trials. To improve the success in obtaining measured data from recorded trials, only the last 36 sec. of each 1-min. trial were processed for time measurements.

An example of the records on which the time measurements were based is shown in Figure 4. In addition to the time line indicating seconds, the two top lines of the record represent the right and left eye movements, respectively, while the bottom lines represent the differentiated velocity records of these two movements.

The reading material used was taken from a collection of readings, each passage of which was ranked according to the Flesch Ease of Reading Scale. Three levels of reading difficulty were used. In the ranked scale used, easy material had the higher numerical value. As expressed by this scale, the three easy passages used had ranks of 85, 84, and 79; the three medium level passages had ranks of 62, 63, and 64; and the three difficult passages had ranks of 33, 41, and 43. Each of

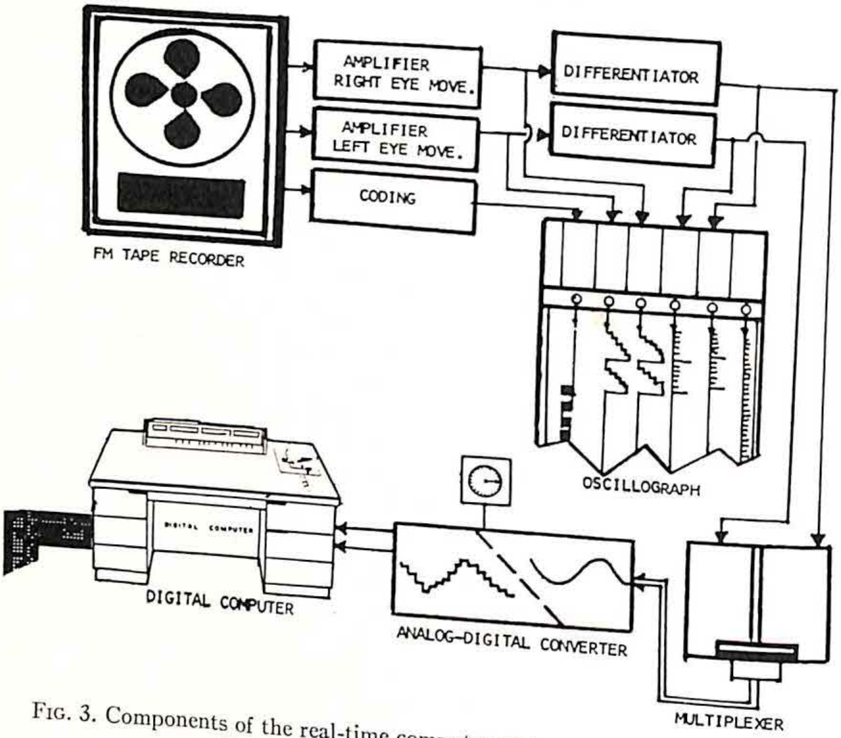


FIG. 3. Components of the real-time computer system used in measuring the time differences between the two eyes.

the nine passages was used for two reading trials of 1-min. each which were run consecutively. A rest period of 30 sec. occurred between trials.

The reading material was typewritten, double spaced, and had a line length of 6 in. The paragraph indentations from the original passages were altered to provide

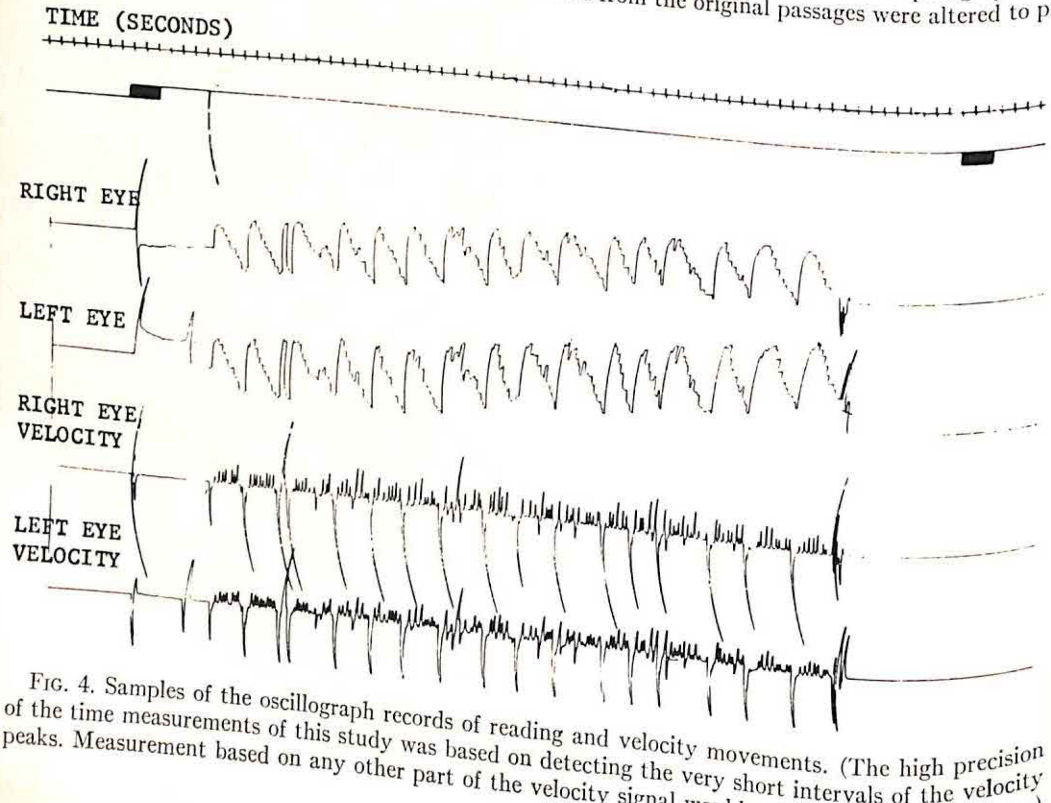


FIG. 4. Samples of the oscillograph records of reading and velocity movements. (The high precision of the time measurements of this study was based on detecting the very short intervals of the velocity peaks. Measurement based on any other part of the velocity signal would not have been as accurate.)

a more uniform reading display, so that each display consisted of two paragraphs 8–10 lines in length. For each trial the reading display included 18–20 lines. The material was typed on yellow paper and displayed on a tel-prompter. This device provided rear lighting of the material and allowed *E* to advance from one reading passage to the next by remote control. The *S* was instructed to read the material silently at his normal rate of reading. In all cases the passage material was completed within the elapsed time of the experimental trial.

Besides varying the difficulty of the reading material, orientation of the reading display was varied by rotating the tel-prompter through 15° to the right and the left. Three *S* were used and were run in a repeated-measures experimental design that encompassed two trials per reading passage, three levels of reading difficulty (easy, medium, and difficult), and three orientations of the reading display (normal, 15° right, and 15° left). In all, there were 18 trials for each *S*. The following scheme summarizes the order of trials (T) for each *S* and also indicates the trials which gave outlying data that was discarded. In the scheme, D_1, D_2, and D_3 stand for low, medium, and high levels of reading difficulty, respectively, and O_1, O_2, and O_3 represent right, normal, and left orientations of the reading display. Asterisks indicate trials from which data could not be recaptured.

$$\begin{array}{cccccccccc}
 & & O_1 & & & O_2 & & & O_3 & \\
S_1 & D_1 & D_2 & D_3 & D_2 & D_1 & D_3 & D_3 & D_2 & D_1 \\
 & T_1\,T_2 & T_1\,T_2 & *\;T_2 & T_1\,T_2 & T_1\;* & T_1\,T_2 & T_1\,T_2 & T_1\;* & T_1\,T_2 \\
 & & O_2 & & & O_1 & & & O_3 & \\
S_2 & D_2 & D_1 & D_3 & D_1 & D_2 & D_3 & D_3 & D_2 & D_1 \\
 & T_1\,T_2 & *\;* & T_1\;* & *\;* & T_1\,T_2 & T_1\,T_2 & T_1\,T_2 & T_1\;* & T_1\,T_2 \\
 & & O_3 & & & O_2 & & & O_1 & \\
S_3 & D_3 & D_2 & D_1 & D_2 & D_1 & D_3 & D_1 & D_2 & D_3 \\
 & T_1\,T_2 & *\;* & T_1\,T_2 & T_1\,T_2 & *\;T_2 & *\;T_2 & T_1\;* & *\;* & T_1\,T_2
\end{array}$$

RESULTS

The results of the experiment consist of time values which represent the extent to which the right or left eye led the other eye during the saccadic reading movements. The nature of the time data obtained may be indicated by the overall means for *S*s and orientation conditions. The grand mean binocular-time difference for all three *S*s was 6.7 msec. which represented a leading time value for the left eye. The precision of this measurement, as determined by generating and processing of artificial electrical inputs to the computer system, was ±1 msec. The mean time values for all three *S*s were 4.0, 9.0, and 7.1 msec. for S_1, S_2, and S_3, respectively, all representing a left eye lead.

The mean time differences for the three *S*s at the three different levels of orientation of the reading display are summarized in Table 1. All of these differences represent left eye leads. The detailed data on which these mean values are based are summarized in Figure 5. These curves plot frequency of measurements of left eye and right eye leads in relation to time in milliseconds. Perfect synchronism or identity of binocular movements is indicated by a zero time value, right eye leads by values given to the right of the zero values, and left eye leads by values to the left of the zero values. The top curve in Figure 5 represents right rotation of the display; the middle curve, normal orientation; and the bottom curve, left rotation.

TABLE 1

MEAN LEFT EYE TIME LEADS FOR THE THREE *S*S AT THE THREE DISPLAY ORIENTATIONS

Subject	Orientation[a]		
	Right	Normal	Left
S_1	4.0	5.6	4.4
S_2	9.9	7.9	9.2
S_3	8.0	5.5	7.2

Note.—All of the values in the table represent left eye leads (in msec.).

[a] For orientation conditions $F = 3.08$, $p > .05$.

The meaning of the differences shown in Figure 5 can be described in still another way in terms of the percentage of left eye and right eye leads that occurred in the measurement. These values are summarized in Table 2 in relation to the three conditions of orientation of the reading display. The table indicates that the left eye led the right eye 71%, 80%, and 86% of the time in the three conditions, whereas the right eye led the left in 8%, 9.5%, and 16 % of the measurements. There were no differences between the two eyes in 6%, 10.5%, and 13% of the time in the three conditions of orientation.

TABLE 2

PERCENTAGE OF LEFT EYE AND RIGHT EYE LEADS AT THE THREE CONDITIONS OF ORIENTATION OF THE READING MATERIAL

Orientation cond.	Left eye lead	Right eye lead	No difference
15° left	80	9.5	10.5
Normal	71	16	13
15° right	86	8	6

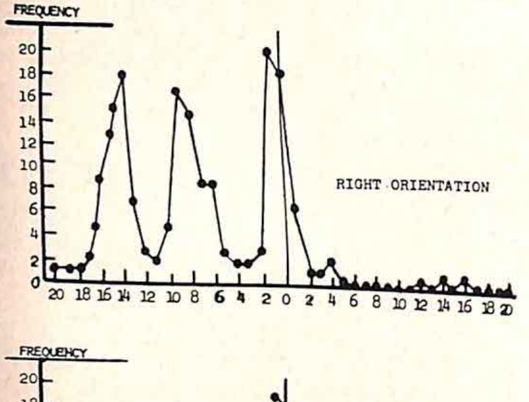


FIG. 5. Distribution of the frequency of time-leads of the left and right eye for different conditions of orientation of the reading material.

To determine if the mean left eye leads for the three orientation conditions were significantly different from one another, Neuman-Keuls (Winer, 1962) range tests were conducted on the data. Results showed that the mean for the left rotation condition was not significantly different from the means for the other two conditions. The mean for right orientation, however, was different from the normal orientation condition.

The actual distribution of time differences between the two eyes was trimodal. As indicated in Figure 5, this trimodal distribution was found for all three conditions of the orientation of the reading display. For the left orientation condition (bottom curve), three modal time values occurred at 1, 9, and 14 msec. For the normal condition, the three modes occurred at 1, 9, and 15 msecs. In the case of the right rotation condition, these modes occurred at 1, 8, and 14 msecs.

The trimodal distribution of eye leads which appeared for the distribution of overall mean values for orientation conditions was found for each of the three *S*s. Figure 6 summarizes the distributions of eye leads for the three *S*s for the least difficult reading passages. These curves show that S_1 differed from the other two *S*s in giving a modal right eye lead at about 5 msecs. This *S* showed a relatively low modal lead value at 16 msecs., and a very high frequency left eye lead at 1 msec. In contrast, S_2 showed a high frequency left eye lead at 14 msec., and a low left lead mode at 1 msec. S_3 showed a somewhat different distribution, with the highest modal frequency at 0 msec.

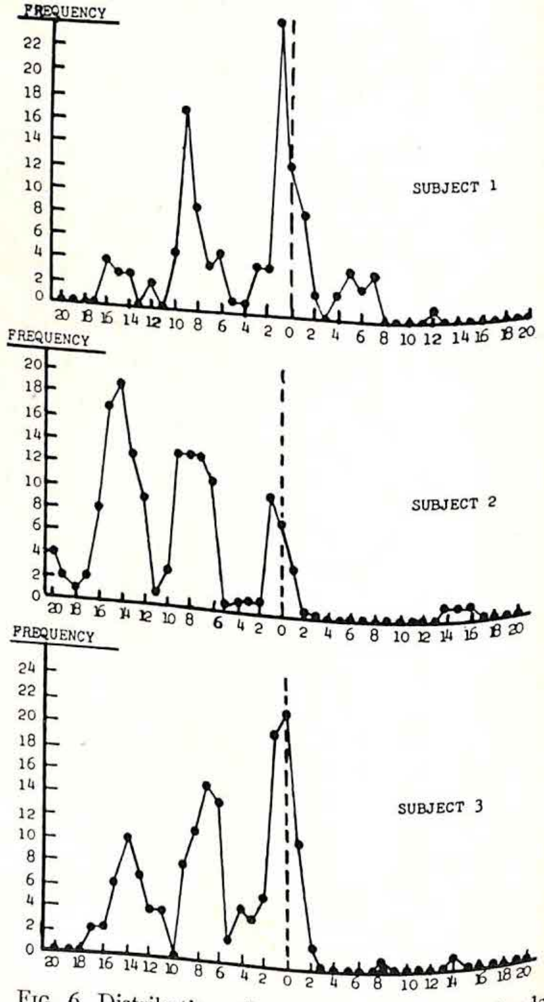


FIG. 6. Distribution of the frequency of time-leads of the left and right eye for different *S*s with the least difficult of the reading passages.

and nearly equal modal frequencies at 14 and 9 msecs. These *S* differences in modal frequencies at different time values were duplicated almost exactly in the other two conditions of difficulty of the reading material.

There are limitations to the data besides the fact that a considerable number of trials could not be completely scored. In addition, the computer procedures detected a considerable number of unpaired eye movements, as defined by the criterion of time pairing of 20 msecs. Table 3 summarizes the number of these unpaired measures for the three *S*s and for the different orientation conditions. Table 3 indicates separately the number of examples in which there was no pairing for a left eye movement and for a right eye movement. It is not known to what extent these failures to find a paired movement represent errors in detection or actual independent movements of one of the eyes. The table shows that *S*s differed in the number of unpaired movements for the left and right eye. S_1 showed the fewest number of unpaired movements and three to four times as many left eye and right eye unpaired movements. S_2 showed roughly four to five times as many unpaired movements as S_1 and approximately 1.3 times as many left eye and right eye unpaired movements. In the case of S_3, the number of unpaired left eye movements was about the same as that for right eye movements. The program used in this study did not make it possible to determine what proportion of these unpaired movements represented actual independent movements, but future research is planned to determine these proportions.

TABLE 3

NUMBER OF UNPAIRED MEASUREMENTS FOR DIFFERENT *S*S AND DIFFERENT ORIENTATIONS OF THE READING MATERIAL

Subject or orientation	Left-eye unpaired	Right-eye unpaired
Subject		
S_1	115	32
S_2	256	193
S_3	202	223
Orientation		
15° left	185	131
Normal	171	129
15° right	207	172

Note.—These frequencies represent about 20% of the total number of paired movements which were scored.

SUMMARY

Real-time laboratory computer methods were developed for high precision measurement of the time differences between the two eyes in reading and other visual tasks. The aim was to determine the validity of the hypothesis that binocular time differences must occur as an aspect of dynamic feedback yoking of the two eyes in coordinate ocular motions. The results found are in keeping with this hypothesis: very small time differences of the general order of 6.6 msecs., with the left eye leading, were found in all three *S*s and under different conditions of difficulty and orientation of the reading material. The precision of the computer measures was determined to be of the order of 1 msec., and was achieved by the special procedures of velocity measurement of eye movements and computer programming used in the study.

The main finding was that not one, but three modal frequencies occurred in the distribution of the time differences between the two eyes, with the left eye leading in each of these cases. These modal values of 1, 8–9, and 14–15 msecs. occurred at different orientations of the reading material. This trimodal distribution of binocular time differences was found consistently in all three *S*s. The basis of this unusual distribution is not known, but possibly may be related to the time discrepancies for the short and long saccadic movements in reading. The results permit the assumption, to be tested in further study, that the longer saccades of reading produced the longer left eye leads, and that the shorter saccades produced the shorter leads.

The finding that the left eye led the right eye in the great majority of the measures made in this study has a limited counterpart in some observations by Hyde (1959). She found that "outward" movements of the eye from a center position were of lower velocity than "inward" movements from an outlying to the center position. The difference was 58° per second. Hyde judged that the eyes are completely conjugate since she could not detect time differences within the range of precision that her film methods permitted (i.e., 1/164

sec.). However, her data suggest that if velocity differences can occur in ocular movements of different direction, binocular time differences also may occur for "inward" and "outward" movements. Since most motions, including reading, may be biased directionally by sidedness, it is possible that in eye motions the left eye may be required to perform "inward" direction movements for a greater period of time than the right eye, and hence could display a greater velocity of motion and a corresponding time lead in motion over the right eye.

The results complement the findings of other studies on delayed eye-movement-retinal feedback in showing that very small time differences govern the coordination of the eyes. Neurally, these time differences may mean that the neurons of the cortex and midbrain, which govern eye motion, may be time specific as well as direction specific in determining the guidance and coordination of the two eyes. The results also give some added substance to previous findings (Smith & Molitor, 1969) that space and form perception and its developmental deficiencies may depend on precise modes of feedback timing in control of saccadic and pursuit movements. This research emphasizes the need for more precise dynamic measurements of ocular behavior in understanding the basis of visual perceptual disabilities.

REFERENCES

ALPERN, M. Muscular mechanisms In H. Davson (Ed.), The Eye. Vol. 3. New York: Academic Press, 1962.

HYDE, J. E. Some characteristics of voluntary human ocular movements in the horizontal plane. *American Journal of Ophthalmology*, 1959, **48**, 85–94.

SMITH, K. U. Cybernetic theory and analysis of learning. In E. Bilodeau (Ed.), *Acquisition of Skill*, New York: Academic Press, 1966.

SMITH, K. U., & HENRY, J. Cybernetic foundations for rehabilitation. *American Journal of Physiological Medicine*, 1967, **46**, 379–465.

SMITH, K. U., & MOLITOR, K. Adaptation to reversed eyemovement-retinal feedback. *Journal of Motor Behavior*, 1969, **1**, 69–87.

SMITH, K. U., PUTZ, V., & MOLITOR, K. Experimental systems analysis of yoked vision. Madison, Wisconsin: University of Wisconsin, Behavioral Cybernetics Laboratory, 1966. (Mimeo)

STROMBERG, E. L. Binocular movements of the eyes in reading. *Journal of General Psychology*, 1938, **18**, 349–355.

VON HELMHOLTZ, H. L. F. Treatise on physiological optics. Rochester, N. Y.: Optical Society of America, 1924–25. 3 vols. (Translation from the 3rd German ed., J. P. C. Southall [Ed.])

WINER, B. Statistical principles of experimental design. New York: McGraw-Hill, 1962.

YARBUS, A. L. Eye movements and vision. New York: Plenum Press, 1967.

(Received December 3, 1969)

JOURNAL OF APPLIED PSYCHOLOGY
MONOGRAPH

Vol. 55, No. 3 June 197

EMPLOYEE REACTIONS TO JOB CHARACTERISTICS[1]

J. RICHARD HACKMAN[2] AND EDWARD E. LAWLER III

Yale University

A conceptual framework specifying the conditions under which jobs will facilitate the development of internal motivation for effective performance was developed and tested. The *S*s were 208 employees of a telephone company who worked on 13 different jobs. Primary independent variables were: (*a*) a measure of strength of desire for the satisfaction of "higher order" needs (e.g., obtaining feelings of accomplishment, personal growth); and (*b*) descriptions of jobs on four core dimensions (variety, autonomy, task identity, feedback). It was predicted and found that when jobs are high on the four core dimensions, employees who are desirous of higher order need satisfaction tend to have high motivation, have high job satisfaction, be absent from work infrequently, and be rated by supervisors as doing high quality work. A number of supplementary analyses were reported, and the implications of the results for future research on job effects and for the design of jobs were discussed.

Researchers and managers alike are increasingly attending to the way jobs are designed as an important factor in determining the motivation, satisfaction, and performance of employees at work. This is not to say that jobs previously have been seen as irrelevant to organizational administration. On the contrary, earlier in this century when scientific management was in its prime, considerable research effort was expended to find ways that jobs could be simplified, specialized, standardized, and routinized. At the same time, industrial psychologists were developing rather complex and sophisticated procedures for describing and analyzing jobs in terms of their simplest components, as a means of evaluating the skill levels required for different jobs. The results of job analyses have been used to establish fair rates of pay, for training purposes, and in personnel selection (see, e.g., Ghiselli & Brown, 1955; Lytle, 1946; Stigers & Reed, 1944). The general expectation of the scientific management approach was that by simplifying jobs, work could be carried out more efficiently; less-skilled employees would be required; the control of management over production would be increased; and, ultimately, organizational profits would be enhanced.

In recent years, numerous scholars have documented a number of unintended and unfortunate consequences of the trend toward work simplification (e.g., Argyris, 1964; Blauner, 1964; Davis, 1957; Friedmann, 1961; Guest, 1955; Herzberg, Mausner, & Snyderman, 1959; Walker, 1950; Walker & Guest, 1952). In brief, it has been shown that simple, routine nonchallenging jobs often lead to high employee dissatisfaction, to increased absenteeism and turnover, and to substantial difficulties in effectively managing employees who work on simplified jobs.[3] The expected increases in profitability from work simplification have not materialized as had been hoped, and the reasons apparently have very much to do with the human problems encountered when jobs are standardized and simplified.

---

[1] This research was supported in part by United States Air Force Office of Scientific Research Grant AF-AFOSR-68-1600, Effects of Task Characteristics on Performance (J. Richard Hackman, principal investigator). The cooperation and assistance of employees and managers in the company where the research was done is gratefully acknowledged, as is the assistance of Ann Garvin and Lloyd Suttle in collecting and analyzing the data. Clayton Alderfer and Chris Argyris provided helpful comments on an earlier draft of this article.

[2] Requests for reprints should be sent to J. Richard Hackman, 2 Hillhouse Ave., Yale University, New Haven, Connecticut 06520.

[3] These observations have not, however, gone unchallenged. See, for example, Kilbridge (1961) and MacKinney, Wernimont, and Galitz (1962).

Partially in response to the above findings, a number of researchers began experimentally enlarging various jobs to determine whether or not worker productivity and satisfaction would increase if jobs were designed so as to be more generally meaningful and challenging to employees. By and large, those job enlargement experiments which have been reported in the literature have been considered successful (see, e.g., Biganne & Stewart, 1963; Conant & Kilbridge, 1965; Davis & Valfer, 1965; Ford, 1969; Kilbridge, 1960; Pelissier, 1965). With few exceptions, however, job enlargement experiments have been case studies and often have lacked appropriate experimental controls. Hulin and Blood (1968) review the research literature on job enlargement in some detail and are especially attentive to possible difficulties in procedure and methodology which may cast doubt on the generality or the validity of the findings reported.

Perhaps equally as disturbing as the uneven level of methodological rigor which has characterized job enlargement studies is the almost total absence of any systematic conceptual or theoretical basis for the studies which have been done. As a result, after dozens of experiments, little cumulative knowledge has been gained regarding the effects and effectiveness of job redesign. Job enlargement experiments, for example, typically have involved a number of simultaneous changes—such as in the amount of variety in the work, the amount of responsibility required, the degree to which working with others is an important part of the enlarged job, etc. Very little is known about which of these (or of other) aspects of the redesigned job are in fact responsible for observed behavioral and attitudinal changes. Further, the generality of job enlargement effects is largely unknown (e.g., whether they are effective only for certain types of workers or whether they are relevant only to certain kinds of jobs). More case studies are not likely to contribute very much to the development of answers to crucial questions such as these. Instead, what appears to be needed are conceptual frameworks which generate testable propositions about how job characteristics affect employees under various circumstances, and empirical research which is designed explicitly to test these propositions. This article proposes one way of conceptualizing the impact of job characteristics on individual work behavior and attitudes. It then reports data which were collected to provide a preliminary test of that conceptualization.

Previous Theory and Research

Some progress toward the development of theory relevant to job design has been made in recent years. The well-known two-factor theory of Herzberg (Herzberg, Mausner, & Snyderman, 1959; Herzberg, 1966), for example, can be used to derive general propositions regarding conditions on the job which will be motivating and satisfying to employees. In particular, the theory suggests that a job should enhance employee motivation to the extent that it provides opportunities for (*a*) achievement, (*b*) recognition, (*c*) responsibility, (*d*) advancement, and (*e*) growth in competence. These principles have given rise to a series of generally successful job enlargement experiments in the American Telephone and Telegraph Company (summarized by Ford, 1969).

Unfortunately, a number of researchers have been unable to provide empirical support for some of the major tenets of the theory from which the principles used in the AT&T studies were derived (e.g., Dunnette, Campbell, & Hakel, 1967; Hinton, 1968; King, 1970), and the general conceptual status of the theory must presently be considered uncertain. Further, the theory has not yet been elaborated to specify how characteristics of workers interact with the presence or absence of the five motivating conditions in determining worker performance and satisfaction. Finally, the theory in its present form does not specify how the presence or absence of the motivating conditions can be measured for existing jobs. This makes it very difficult to test the theory and to generate unambiguous predictions from it about the effects of specific changes which may be contemplated for existing jobs.

The problem of measuring job characteristics has been explicitly and carefully dealt with by Turner and Lawrence (1965). These authors developed operational measures of six "requisite task attributes" which, on the basis of a review of existing literature and an a priori conceptual framework, were predicted to be positively related to worker satisfaction and attendance. The six attributes are: (*a*) variety, (*b*) autonomy, (*c*) required interaction, (*d*) optional interaction, (*e*) knowledge and skill required, and (*f*) responsibility. Scores on each attribute for each of 47 different jobs were obtained from field observations and interviews by the researchers, and precise specification of how scores on each of the attributes is determined is provided. Examination of the relationships among the six requisite task attributes for the 47 jobs revealed that the attributes were very closely related to one another. Therefore, Turner and Lawrence developed a summary measure called the Requisite Task Attribute Index (RTA Index) by formulating a linear combination of the six separately measured attributes. This summary index was then used in ascertaining the relationships between the attributes of the jobs and worker job satisfaction and attendance.

The authors' expectation that employees working on jobs which were high on the RTA Index would have higher job satisfaction and lower absenteeism was not fully supported. Instead, it appeared that the predicted relationship between the RTA Index and employee reactions held only for workers from factories located in small towns. Workers in urban settings reported less satisfaction with their jobs when the jobs were high on the RTA Index, and the RTA Index was unrelated to absenteeism for urban workers. Turner and Lawrence (1965) argued that the obtained differences in reactions to good (i.e., high RTA Index) jobs were substantially moderated by differences in the cultural backgrounds of employees.

Blood and Hulin (1967) and Hulin and Blood (1968) provide additional data on the importance of subcultural factors in determining worker responses to the makeup of their jobs. These authors hypothesize that an important moderating factor is alienation from the traditional work norms which characterize the middle class. When employees hold traditional values regarding the value of work and achievement in work settings (as would be expected of the employees in small town factories in the Turner and Lawrence study), more complex jobs should be responded to positively. When employees are alienated from these norms (as might be expected of urban workers), more complex jobs should be responded to negatively. Blood and Hulin (1967) provide data supporting this general proposition and propose a three-dimensional response surface (Hulin & Blood, 1968) which specifies the expected interrelationships among worker alienation, job level, and satisfaction with work.

The data of Turner and Lawrence (1965) and of Blood and Hulin (1967) are sufficiently compelling that the generality of the strong proposition that enlarged jobs (in the sense of being high, e.g., on the dimensions which make up the RTA Index) lead to improved satisfaction, attendance, and/or performance on the job must be called into question. Instead, it appears that certain characteristics of the employees themselves must be taken into account simultaneously with the characteristics of their jobs in order to generate valid predictions about the behavioral and affective responses of employees at work.

If the above conclusion is accepted, thorny questions then arise about how the relevant differences among employees are to be conceptualized and measured. Both Turner and Lawrence (1965) and Hulin and Blood (1968) choose to deal with individual differences on a subcultural or sociological level (i.e., in terms of differences between town and city workers or in terms of the alienation of city workers from middle-class work norms).

An alternative strategy would be to attempt to conceptualize and measure the relevant individual differences directly at the individual level of analysis. The town-city conceptualization assumes a substantial homogeneity of worker characteristics and response tendencies for employees within the two cultural settings. To the extent that there are substantial individual differences among town workers and among city workers, an attempt to measure relevant individual differences directly at the individual level would seem to have considerable merit. The difficulty in implementing this alternative approach, of course, is that it requires prior specification on a conceptual level of what specific differences among people are responsible for the results reported by Turner and Lawrence (1965) and Blood and Hulin (1967) (i.e., what it is about people that moderates the way they react to their jobs). In the following paragraphs, we will propose such a conceptualization, and derive from it a number of predictions about the effects of job characteristics on employee satisfaction and motivation.

Jobs and Individuals: A Conceptual Framework

The present conceptualization of the interaction between job characteristics and individual differences is based primarily on the expectancy theory of motivation, as formulated by Lewin (1938) and Tolman (1959), and as applied to work settings by Vroom (1964), Porter and Lawler (1968), and others. In particular, five propositions based on expectancy theory are suggested below, which address the specific problem of how employee motivation can be enhanced through the design of jobs.

1. To the extent that an individual believes that he can obtain an outcome he values by engaging in some particular behavior or class of behaviors, the likelihood that he will actually engage in that behavior is enhanced. Relevant valued outcomes can be both intrinsic (e.g., feelings of accomplishment or of personal worth) and extrinsic (e.g., material goods); the only requirement is that the outcomes be valued by the individual. When an individual anticipates obtaining some valued outcome as a result of a contemplated action or course of action, that outcome may be termed an incentive for engaging in the action.

2. Outcomes are valued by individuals to the extent that they satisfy the physiological or psychological needs of the individual, or to the extent that they lead to other outcomes which satisfy such needs or are expected by the individual to do so. Such need satisfaction need not, of course, be in the objective best interest of the individual. People frequently strive for satisfying states of affairs which are quite inconsistent with their long-term well-being (Locke, 1969). Nevertheless, if an outcome is not somehow linked to satisfaction, it will not continue to be valued and therefore cannot continue to serve as an incentive.

3. Thus, to the extent that conditions at work can be arranged so that employees can satisfy their own needs best by working effectively toward organizational goals, employees will in fact tend to work hard toward the achievement of these goals (McGregor, 1960).

4. Most lower level needs (e.g., physical well-being, security) can be (and often are) reasonably well satisfied for individuals in contemporary society on a continuing basis and, therefore, will not serve as motivational incentives except under unusual circumstances. This is not the case, however, for certain higher order needs (e.g., needs for personal growth and development or for feelings of worthwhile accomplishment). A person may experience higher order need satisfaction on a continuing basis without the strength of desire for additional satisfaction of these needs diminishing. Indeed, it may be that additional satisfaction of higher order needs actually increases their strength (Alderfer, 1969). This is an important possibility since it suggests that the opportunity for the development of continuing (and possibly even increasing) motivation is much more a reality when higher order needs are engaged than is the case for more easily satisfied lower order needs. There is, of course, a major cost associated with any motivational approach in which higher order needs are central: Not all employees can or will respond to opportunities for the satisfaction of higher order needs, and thus motivational approaches based on these needs cannot be applied indiscriminantly. Maslow (1943, 1954) and Alderfer (1969, 1971) discuss in much more complete detail the nature of higher order needs and their motivational implications.

5. Individuals who are capable of higher order need satisfaction will in fact experience such satisfaction when they learn that they have, as a result of their own efforts, accomplished something that they personally believe is worthwhile or meaningful (see Argyris, 1964; Lewin, Dembo, Festinger, & Sears, 1944). Specifically, individuals who desire higher order need satisfactions should be most likely to obtain them when they work effectively on meaningful jobs which provide feedback on the adequacy of their personal work activities. To establish conditions for internal work moti-

ation, then, it appears that a job must: (*a*) llow workers to feel personally responsible or an identifiable and meaningful portion of he work, (*b*) provide work outcomes which re intrinsically meaningful or otherwise exerienced as worthwhile, and (*c*) provide feedack about performance effectiveness. The arder and better an individual works on such job, the more opportunities he will have to xperience higher order need satisfactions and he more incentive there can be for continued ffective performance. Higher order need satisfactions, therefore, are seen both as (*a*) a result of (rather than a determinant of) effective performance (Lawler & Porter, 1967), and (*b*) an incentive for continued efforts to perform effectively.[4]

The five propositions outlined above lead to the conclusion that it may be possible under specifiable conditions simultaneously to achieve high employee satisfaction and high employee effort toward organizational goals. Specifically, the long term congruence of high satisfaction and high effort is seen as depending upon (*a*) the existence of employee desires for higher order need satisfaction and (*b*) conditions on the job such that working hard and effectively toward organizational goals will bring about satisfaction of these needs.

Characteristics of Motivating Jobs

The three general job characteristics identified above as central in developing a congruence between individual need satisfaction and organizational goal achievement must be describable in more measurable terms if the validity of the conceptualization proposed here is to be tested. In the following paragraphs, therefore, each of the three general characteristics are examined in somewhat more detail. In addition, it will be proposed that four of the requisite task attributes proposed by Turner and Lawrence (1965) are likely to be useful as measures of the three general job characteristics.

1. The job must allow a worker to feel personally responsible for a meaningful portion of his work. What is accomplished must be through the individual's own efforts. He must realize that the work he does is his own. And he must believe that he personally is responsible for whatever successes and failures occur as a result of his work. Only if what is accomplished is seen as one's own can an individual experience a feeling of personal success and a gain in self-esteem. This does not mean, of course, that feelings of personal responsibility for work outcomes cannot occur in team projects; all that is required is for team members to feel that their own efforts are important in accomplishing the task at hand.

The autonomy dimension, as specified by Turner and Lawrence (1965), would seem to tap the degree to which workers feel personal responsibility for their work. In jobs high on measured autonomy, workers will tend to feel that they own the outcomes of their work; in jobs low on autonomy, a worker may more often feel that successes and failures on the job are more often due to the good work (or to the incompetence) of other workers or of his supervisor.[5]

2. The job must provide outcomes which are intrinsically meaningful or otherwise experienced as worthwhile to the individual. If a worker feels that the results of his efforts are not very important, it is unlikely that he will feel especially good if he works effectively. Instead, he must achieve something that he personally feels is worthwhile and important if he is to be able to experience positive feelings about himself as a result of his efforts. It clearly is not possible to indicate for people in general what kinds of job characteristics will be likely to provide outcomes seen as meaningful and worthwhile; people differ too much in the kinds of things they value for any statement of such generality to be made. It is possible, however, to provide some such

---

[4] It should be noted that only higher order satisfactions are predicted to be increased by effective work on a job with the characteristics outlined above; other satisfactions (e.g., pay satisfaction, satisfaction with supervision) may not be affected.

[5] Having high autonomy on the job does not, of course, necessarily imply that one will have major control over the work outcomes achieved. There may be a number of factors in the work environment which affect the nature of work outcomes, over which the worker has little meaningful control. For example, a football quarterback has high autonomy in selecting plays but only a moderate level of personal control over the outcomes obtained from execution of the plays. Thus, work autonomy is probably best viewed as a necessary but not sufficient condition for feeling personal responsibility for work outcomes.

specifications for individuals who have high desires, for higher order need satisfaction and, of course, these are the individuals to whom the present conceptualization is intended to apply.

There are at least two ways that work can come to be experienced as meaningful for employees with relatively high desires for higher order need satisfaction. The first is for the job to be a sufficiently whole piece of work that the worker can perceive that he has produced or accomplished something of consequence. In terms of a Turner and Lawrence task attribute, this would be expected to be the case when a job is high on task identity. According to Turner and Lawrence (1965, p. 157), jobs high on task identity are characterized by (*a*) a very clear cycle of perceived closure—the job provides a distinct sense of the beginning and ending of a transformation process, (*b*) high visibility of the transformation to the worker, (*c*) high visibility of the transformation in the finished product, and (*d*) a transformation of considerable magnitude. For a worker who has high needs for developing and using his competence, a job with such characteristics generally would be expected to be experienced as highly meaningful and worthwhile.

In addition, the experienced meaningfulness of work may be enhanced when a job provides a worker with the opportunity to accomplish something by using skills and abilities which he personally values. For example, a strongly motivated duffer feels good when he hits a solid tee shot, even though the broader significance of this event is doubtful. His golfing skills are on the line when he steps to the tee; those skills are important to him; he performs well—and that, in itself, is enough.

Jobs high on the Turner and Lawrence (1965) dimension of variety would be expected to provide opportunities for workers to experience this kind of meaningfulness on the job, since high variety jobs typically tap a number of different skills which may be important to the employee. Thus, working on high variety jobs may become personally meaningful to some employees through a process very analogous to that which makes golf meaningful to the duffer. It should be noted, however, that only variety which does in fact challenge the worker will be expected to be experienced as meaningful to workers with desires for higher order need satisfaction; screwing many different sizes of nuts on many different colors of bolts, if this could be considered variety, would not be expected to be experienced as meaningful.[6]

To summarize, it may be that jobs can come to be experienced as meaningful to employees to the extent that they involve doing a whole piece of work of some significance (i.e., have high task identity) and, at the same time, to the extent that they give employees the chance to use their valued skills and abilities (i.e., to be challenged) in doing the work. In many cases the latter condition may be met on jobs which have high variety.

3. The job must provide feedback about what is accomplished. Even if the two general conditions discussed above are met, an employee cannot experience higher order need satisfaction when he performs effectively unless he obtains some kind of feedback about how he is doing. Such feedback may come from doing the task itself (e.g., when a telephone operator successfully completes a long distance person-to-person call), but performance feedback also may come from some other person—an esteemed co-worker, a supervisor, etc. The crucial condition is that feedback be present in a form that is believable to the worker, so that a realistic basis exists for the satisfaction (or frustration) of higher order needs.

It should be emphasized that, for all of the job characteristics discussed above, it is not their objective state which affects employee attitudes and behavior, but rather how they are experienced by the employees. Regardless of the amount of feedback (or variety, or autonomy, or task identity) a worker really has in his work, it is how much *he perceives that he has* which will affect his reactions to the job. Objective job characteristics are important because they do affect the perceptions and experiences of employees. But there are often substantial differences between objective job characteristics and how they are perceived by employees, and it is dangerous to assume that simply because the objective characteristics of a job have been measured (or changed) that the way that job is experienced by employees has been dealt with as well.

---

[6] It is also possible, of course, for a job to have too much variety. Activation theory (e.g., Scott, 1966) suggests that when variety is too high, employees may experience a general state of muscular and mental hypertension which can greatly handicap performance effectiveness. In addition, Hall and Lawler (1970) found that among research scientists, high job variety can be associated with low job satisfaction, apparently because jobs with high variety also tended to be low in task identity and feedback.

In summary, then, it has been argued that the characteristics of jobs can establish conditions which will enhance the intrinsic motivation of workers who desire higher order need satisfaction. In particular, it has been suggested, in terms of a subset of the Turner and Lawrence (1965) dimensions, that such individuals will be able to obtain meaningful personal satisfaction when they perform well on jobs which they experience as high on variety, autonomy, task identity, and feedback. Further, the harder and better an individual performs on a job which is perceived as high on these dimensions (hereafter will be called the four core dimensions), the more satisfaction he is likely to feel.

Strategy of the Present Research

The conceptualization presented above provides the basis for the present empirical research on the relationships among job characteristics; individual differences in need strength; and employee motivation, satisfaction, performance, and absenteeism on the job. In particular, the research to be reported here follows the strategy steps listed below:

1. Measures of the following six job dimensions were developed: (*a*) variety, the degree to which a job requires employees to perform a wide range of operations in their work and/or the degree to which employees must use a variety of equipment and procedures in their work; (*b*) autonomy, the extent to which employees have a major say in scheduling their work, selecting the equipment they will use, and deciding on procedures to be followed; (*c*) task identity, the extent to which employees do an entire or whole piece of work and can clearly identify the result of their efforts; (*d*) feedback, the degree to which employees receive information as they are working which reveals how well they are performing on the job; (*e*) dealing with others, the degree to which a job requires employees to deal with other people (either customers, other company employees, or both) to complete the work; (*f*) friendship opportunities, the degree to which a job allows employees to talk with one another on the job and to establish informal relationships with other employees at work.

The latter two dimensions were included to permit exploration of the impact of the interpersonal characteristics of job design. These dimensions were adapted with very minor revision from the task attributes "required interaction" and "optional interaction" proposed by Turner and Lawrence (1965). They are not, however, directly relevant to the conceptualization about job-based work motivation proposed above, and no specific predictions regarding them were made. Thirteen different jobs were described on these six dimensions by the researchers, by employees who worked on the jobs, and by members of the management of the telephone company in which the research was carried out.

2. A measure was developed which was expected, on an a priori basis, to reflect the level of employee desire for the satisfaction of higher order needs.

3. Based on the mean scores of the employees on the measure of need strength, predictions were made regarding the expected relationships between the job characteristics as measured by the four core dimensions and the dependent variables: satisfaction, performance, and absenteeism. Relevant data were collected from 208 employees, and correlations between each of the four core dimensions and each of the dependent measures were computed.

4. The theory outlined above indicates that how a job is experienced or perceived by an individual employee should determine his reactions to it, rather than the objective characteristics of the job. This possibility can be examined by analyzing the relationship between the characteristics of a single job (as they are perceived by employees) and the behavioral and attitudinal reactions of individuals who hold that job. By restricting the scope of the analysis to individuals working on the same job, it is possible to rule out objective differences in jobs as an explanation for

obtained empirical relationships—and, thereby, to address the possibility that the perceived characteristics of jobs affect employees in the same general fashion as do objectively measured characteristics. Therefore, relationships between perceived job characteristics and the dependent variables of interest were computed separately for employees who worked on each of the 13 objectively different jobs included in the study, and these relationships were compared to those obtained in the overall analyses involving all *S*s and all 13 jobs.

5. The theory implies that satisfaction, performance, and attendance should be highest when all four of the core dimensions are present. That is, being high on, say, variety and task identity but low on autonomy and feedback should not provide the conditions that are necessary for positive behavioral and affective responses. Under such circumstances, a worker would be expected to have neither a sense of personal ownership of his work activities and outcomes (low autonomy) nor a trustworthy gauge of the adequacy of his performance (low feedback). The conditions for the development of internal work motivation specified earlier would therefore not be met. The importance of having all four core dimensions present was tested by comparing the dependent variable scores of employees who saw their jobs as moderately high on all four core dimensions both with the scores of employees who saw their jobs as high on some dimensions and low on others, and with the scores of employees who saw their jobs as moderately low on all four dimensions.

6. The theory states that individual differences in desire for higher order need satisfactions should moderate the relationships between job characteristics and the dependent variables. In order to test this possibility, relationships were computed separately and compared for the third of the employees highest on desire for higher order need satisfaction and for those employees lowest on desire for higher order need satisfaction.

7. Finally, exploratory analyses were made of the relationships between the two interpersonal job dimensions (dealing with others, and friendship opportunities) and the dependent variables.

METHOD

Research Setting and Subjects

The research was carried out in an eastern telephone company, and focused on employees who worked on 13 different jobs in the plant and traffic departments of the company. The jobs were selected so as to include (*a*) several varieties of operators, installers, central office repairmen, and cable splicers; (*b*) to range widely in complexity and in the level of employee skill required; and (*c*) to be located in both rural and urban settings.

Data were collected from 208 employees and 62 supervisors. Employees in the traffic department (about one-third of the sample) were female; all plant department employees were male.

Procedure

All data were collected on site at each of the 13 job locations. One to three days were spent at each location, and all data were collected from each location before moving on to another location. Data were collected over a period of about 9 mo. At each location, the following five procedural steps were followed (although the order of the steps sometimes varied because of local circumstances).

1. Local second- or third-level management was visited to obtain permission to collect data from employees working on a particular job. When permission was secured (it was never denied), the managers were interviewed about the general nature of the job as they perceived it.

2. First-level supervisors on the local job were interviewed about the nature of the job and employees' reactions to it and were given a questionnaire which tapped the supervisors' perceptions of the employee job in a format similar to that used for obtaining the employees' own job perceptions.

3. Employees working on the job were observed and interviewed informally. These observations and interviews were conducted by two researchers and continued until the researchers felt that they were no longer obtaining substantial new information about the job. Observation typically lasted approximately one working day. Ratings of the job characteristics by the researchers were made on the basis of these observations and interviews.

4. A questionnaire was administered to a sample of 15–20 employees on the job. The questionnaire took between $\frac{1}{2}$–1 hr. to complete and usually was administered to employees in groups of three or four. The general nature of the research was explained to each group of employees before they began work on the questionnaire, although the hypotheses of the study and the dependent variables to be analyzed were not mentioned. It was emphasized to each individual that participation was voluntary, and a few individuals did decline to participate. In addition, employees were told that putting their names on the questionnaires, while desirable for research purposes, also was voluntary. About 10% of the employees who participated declined to give their names.

5. Ratings of the performance of those employees who had taken the questionnaire and provided their names were made by first-line supervisors. Absence data were collected later from company records.

Instruments and Measures

Job descriptions. Each of the 13 jobs examined in the research was described in terms of the six general dimensions described earlier (variety, autonomy, task identity, feedback, dealing with others, and friendship opportunities). The jobs were described on all six dimensions using four different methods: (*a*) a sample of employees on the job completed several questionnaire scales relevant to the six dimensions; (*b*) first- and second-level management on the job completed identical questionnaire scales; (*c*) the researchers, after observing the job and interviewing employees and supervisors, completed similar scales; and (*d*) the researchers utilized a set of objective coding procedures adapted from Turner and Lawrence (1965) to obtain a set of nonsubjective assessments of each job.

The job descriptions completed by employees and supervisors involved three separate questionnaire items for five of the six dimensions; two items were used for the dimension "dealing with others." One section of the questionnaire involved making direct ratings of the job on six 7-point scales. The scales used in this section of the employee questionnaire, together with the labels attached to each of the scale end points and the midpoint, are listed below.

a. How much *variety* is there in your job?
Very little; I do pretty much the same things over and over and use the same pieces of equipment and procedures almost all the time. (Scored 1)
Moderate variety (Scored 4)
Very much; I do many different things and use a wide variety of equipment and procedures. (Scored 7)

b. How much *autonomy* do you have on your job; how much are you left on your own to do your own work?
Very little; I have almost no "say" about scheduling my work; the work and the procedures are all laid out for me in detail.
Moderate autonomy; I make some of the decisions about my work, but many of them are made for me.
Very much; I have almost all of the "say" about the scheduling of my own work; I alone decide what procedures will be used.

c. To what extent do you *do a "whole" piece of work* (as opposed to doing part of a job which is finished by some other employee)?
I do one small part of a job; there are many others who do other parts of the job; I may not see the final result.
I do a moderate size "chunk" of the work; there are others involved too, but my contribution is clear.
I do an entire piece of work; I do the job from start to finish, and what is done is clearly "mine."

d. To what extent do you *find out how well you are doing* on the job as you are working?
Very little; I often work for long stretches without finding out how I am doing.
Moderately; sometimes I know how I am doing and other times I do not.
Very much; I get almost constant "feedback" on my performance as I work.

e. To what extent do you have the *opportunity to talk informally* with other employees while at work. (That is, is your job arranged so that you can chat with other workers while on the job—even though the job does not require you to talk to these people?)
Very little; there is almost no chance to talk to other employees except about "business."
Moderately; there is some chance to talk, but you may have to arrange it ahead of time.
Very much; there is almost always an opportunity to talk with other employees about non-business topics.

f. To what extent is *dealing with other people* a part of your job?
Very little; working with other people is *not* a very important part of my job.
Moderately; I have to deal with some other people, but this is not a major part of my job.
Very much; probably the single most important part of my job is working with other people.

A second section of the questionnaire asked employees on 7-point scales how much of the various job attributes was actually present on their jobs. These scales ranged from "none or a minimum amount" (scored 1) through "a moderate amount" (scored 4) to "a maximum amount" (scored 7). Scales which were used in this section of the questionnaire include:

(*a*) For variety: The amount of variety in my job.
The opportunity to do a number of different things.

(*b*) For autonomy: The opportunity for independent thought and action.
The freedom to do pretty much what I want on my job.

(*c*) For task identity: The opportunity to do a job from the beginning to end (i.e., the chance to do a whole job).
The opportunity to complete work I start.

(*d*) For feedback: The opportunity to find out how well I am doing in my job.
The feeling that I know whether I am performing my job well or poorly.

(*e*) For dealing with others: The opportunity, in my job, to give help to other people.

(*f*) For friendship opportunities: The opportunity in my job to get to know other people.
The opportunity to develop close friendships in my job.

The items described above were randomly spread throughout a 23-item section of the questionnaire which asked about the presence or absence of a number of heterogeneous aspects of the work setting.

Scores on all items describing each of the six job dimensions were averaged to arrive at a set of six summary scores from the employee ratings described

TABLE 1

ESTIMATED RELIABILITIES OF THE SIX DESCRIPTIVE DIMENSIONS

Dimension	Source of ratings		
	Employees	Supervisors	Turner-Lawrence
Variety	.90	.91	.86
Autonomy	.77	.68	.89
Task identity	.77	.86	.95
Feedback	.75	.75	.97
Dealing with others	.47	.17	.88
Friendship opportunities	.43	.42	.92

Note.—Reported reliabilities for employee and supervisor ratings are estimated internal consistencies. Each dimension score for these ratings consists of the average of three questionnaire items, except for "dealing with others" which is the average of two items. The reliability reported for the Turner-Lawrence procedure is the estimated reliability of the average of two judges, adjusted by Spearman-Brown procedures. Ratings by the researchers (the fourth procedure used) were made collaboratively and reliability data are not available.

above. The questionnaire items given to supervisors were identical in every respect, except that supervisors were asked to describe "the job you supervise" rather than "your job" as was the case for the employees.

Internal consistency reliabilities for the items making up each of the six job dimension scores are presented separately for employees and supervisors in Table 1. Also included are interjudge reliabilities for the researchers' assessments of the jobs. With the exception of supervisors' data on the dimension "dealing with others," the reliabilities are adequate.

Comparison of the scores obtained for each job by the four rating procedures reveals high agreement on all but one of the dimensions. These correlations are presented in Table 2. There is no agreement about which jobs are high and which jobs are low on the feedback dimension. One possible source of confusion regarding this dimension is that some feedback is provided directly by the job as an employee works (e.g., an installer successfully installing and testing a telephone), while additional feedback may be provided by supervision (e.g., a foreman telling an installer that he has taken too long to do a particular job). It may be that employees, supervisors, and researchers were not attending to the same aspect of the feedback process in making their judgments, thereby lowering the level of agreement.

In addition to the problem with the feedback dimension, there is a tendency for correlations involving supervisors to be lower than others. This is especially true for the two dimensions involving interpersonal relationships: dealing with others, and friendship opportunities. Examination of the mean scores of employees and supervisors on these two jobs reveals no consistent differences. Supervisors did not see the jobs as having either consistently more or consistently less of an interpersonal component than employees—they simply did not agree well with the employees (or with the researchers) about the extent of interpersonal activities relevant to the job. Part of the reason for this lack of agreement may be the low internal consistency reliability of supervisors on the dimension "dealing with others." In addition, on-site observations indicated that, by and large, supervisors were considerably more attentive to and concerned with the technical and production aspects of the jobs they supervised than with interpersonal issues. This tendency may partially account for both the unreliability of the supervisors on the "dealing with others" dimension and the low level of agreement between supervisors and the other raters on the two interpersonal dimensions.

Table 2 shows that employees themselves provide judgments of the characteristics of their jobs which, in general, agree quite well with those made by outsiders and (with the exceptions noted above) by their supervisors. It is not, of course, possible to demonstrate conclusively that the employee judgments are objectively accurate, because no unambiguous standard of accuracy is available. Nevertheless, the strong convergence of the employee judgments with the assess-

TABLE 2

CORRELATIONS AMONG THE FOUR RATING PROCEDURES

Correlation between	Job dimensions					
	Variety	Autonomy	Task identity	Feedback	Dealing with others	Friendship opportunities
Employees & supervisors	.87*	.85*	.65*	.09	.31	.49*
Employees & researchers	.94*	.94*	.79*	−.22	.91*	.73*
Employees & Turner-Lawrence	.89*	.72*	.63*	−.66	.95*	.52*
Supervisors & researchers	.94*	.79*	.63*	−.15	.37	.38
Supervisors & Turner-Lawrence	.84*	.49*	.62*	.23	.26	.35
Researchers & Turner-Lawrence	.90*	.80*	.69*	.16	.93*	.67*

Note.—$N = 13$ jobs.
* $p < .05$ (one-tailed test).

TABLE 3

CORRELATIONS AMONG THE SIX DESCRIPTIVE DIMENSIONS: EMPLOYEE RATINGS

Dimensions	Variety	Autonomy	Task identity	Feedback	Dealing with others	Friendship opportunities
Variety	—					
Autonomy	.67*	—				
Task identity	.17	.24*	—			
Feedback	.11	.06	.12	—		
Dealing with others	.05	−.05	.19*	.26*	—	
Friendship opportunities	.41*	.21*	.07	.32*	.13	—

Note.—N = 208 employees.
* $p < .01$ (two-tailed test).

ments obtained from the researchers, from supervisors, and from the Turner-Lawrence procedures does suggest that the employees were able to provide generally nondistorted descriptions of the characteristics of their jobs. In most of the analyses reported in this article, employee judgments are used as the primary measures of job characteristics. This is appropriate and necessary, given that it was argued earlier that employees' perceptions of their jobs (rather than objective job characteristics) are causal in affecting the reactions of employees to their work. Fortunately, the data in Table 2 establish that employee perceptions have convergent validity as descriptors of jobs—and, therefore, that they are probably reasonably well-grounded in reality.

The relationships among the six job dimensions themselves are presented in Table 3. Although there is some tendency for the six dimensions to be positively related to one another, only two of the correlations are of substantial magnitude: jobs seen as having high variety also are seen as being high in autonomy and in friendship opportunities. The level of interrelationship among the six dimensions as measured in the present research is lower than that reported by Turner and Lawrence (1965), and does not mitigate against use of the six dimensions separately as descriptors of job characteristics.

Table 4 presents a listing of the 13 jobs studied, the number of employees on each job who participated in the research, and the mean employee ratings of the characteristics of each job.[7]

Individual need strength. To obtain a measure of the degree to which *S*s were desirous of obtaining higher order need satisfactions from their work, a questionnaire was administered which asked how much of various opportunities and attributes the employees "would like" to have on their job. In a space beside each item, employees entered a number ranging from 1 (would like to have none or a minimum amount) to 7 (would like to have a maximum amount). Content of the questionnaire ranged widely, and included items relevant to pay, promotion, security, working conditions, peer relationships, and supervisory relationships. Twelve of the items were judged on an a priori basis to measure desire for higher order need satisfactions. These items are:

(*a*) The opportunity for personal growth and development on my job.
(*b*) The opportunity for independent thought and action on my job.
(*c*) The opportunity to find out how I am doing.
(*d*) The opportunity to complete work I start.
(*e*) The opportunity to do challenging work.
(*f*) The feeling that I know whether I am performing my job well or poorly.
(*g*) The opportunity to do a number of different things.
(*h*) The opportunity to do a job from the beginning to the end (that is, the chance to do a whole job).
(*i*) The freedom to do pretty much what I want on my job.
(*j*) The amount of variety in my job.
(*k*) The feeling of worthwhile accomplishment in my job.
(*l*) The opportunity, in my job, for participation in the determination of methods, procedures, and goals.

Scores of each employee on the 12 items were summed to obtain an overall measure of the level of

---

in the telephone company and with employee sex (see Table 4). It was concluded on these grounds that the sample of 13 jobs was not adequate for analysis of relations between objective job descriptions and mean dependent variable scores using jobs as observations.

[7] Descriptions of the 13 jobs on Turner-Lawrence dimensions are not included in Table 4 for two reasons. First, the Turner-Lawrence procedures were modified somewhat for use in the present study, and therefore the scores obtained are not directly comparable to those reported by Turner and Lawrence (1965) for their sample of 47 jobs. Second, the primary use of the Turner-Lawrence scores in the present research was to demonstrate the convergence of employee assessments of their jobs with more objectively based job descriptions (see Table 2). Substantive analyses using the modified Turner-Lawrence scores are not reported, since the scores of specific jobs on the six dimensions are confounded both with traffic versus plant functions

TABLE 4

Summary of the Characteristics of the Jobs Studied

Job description	Sex	N	$\bar{X}$ employee ratings of job characteristics					
			Variety	Autonomy	Task identity	Feedback	Dealing with others	Friendship opportunities
Director assistance (DA) operators	F	16	3.08	3.64	5.66	4.88	6.06	4.13
Toll operators	F	18	2.87	3.06	5.15	4.80	6.03	3.69
Combined DA & toll operators	F	17	3.97	3.29	5.87	5.59	6.32	5.02
Traffic service position (TSP) operators	F	18	2.78	2.91	5.06	5.09	6.14	4.23
TSP operators with "enriched" jobs	F	19	2.94	2.96	5.26	5.08	5.90	4.05
Tip & ring (residence) installers	M	18	4.78	5.23	6.38	4.41	5.23	4.14
Key (commercial) installers	M	19	5.72	5.06	6.26	4.87	5.60	4.85
Private branch exchange installers	M	15	6.31	5.61	5.25	4.11	5.07	4.87
Cable splicers	M	16	5.14	4.67	5.04	5.12	4.59	4.67
Combined tip & ring & key installers	M	17	4.94	5.09	6.11	4.23	4.85	4.45
Central office framemen	M	8	4.50	3.89	4.42	5.43	5.12	5.44
Central office repairmen: Step equipment	M	11	5.30	4.84	5.32	4.30	4.77	4.70
Central office repairmen: Crossbar equipment	M	16	5.91	4.92	5.03	5.15	5.47	5.19

higher order need strength. Internal consistency reliability of the 12-item scale was .89.

Employee reactions to their jobs. Four types of employee reactions to their jobs were obtained from the questionnaires employees completed, for use as dependent variable measures. The four types of reactions are described separately below.

1. Experienced work motivation: measures were derived from the questionnaire to reflect both the amount of intrinsic motivation employees experienced on the job and the focus of their motivation. The amount of intrinsic motivation was assessed by three questionnaire items: (*a*) I feel a great sense of personal satisfaction when I do my job well. (*b*) Doing my job well increases my feeling of self-esteem. (*c*) I feel bad when I do my job poorly. The items were placed on 7-point scales ranging from "strongly disagree" (scored 1) to "strongly agree" (scored 7). Scores on the three items were averaged to yield a score indicative of the degree to which an employee experiences positive affective outcomes when he performs well, and negative affective outcomes when he performs poorly (Lawler & Hall, 1970). Internal consistency reliability of this scale was .72.

The focus of employee motivation was tapped by three questionnaire items which addressed the kinds of internal pressures for performance which were experienced on the job. Three items were used, which dealt with three different (but not mutually exclusive) foci of work motivation: (*a*) Being personally responsible for what you do; checking your own work. (*b*) Producing a large quantity of work. (*c*) Doing high quality work. Each item was on a separate 7-point scale, which ranged from high experienced pressure (e.g., "It is *extremely* important for a worker on this job to do high quality work") to low experienced pressure (e.g., "It is not especially important for a worker on this job to do high quality work").

2. Job involvement: three questionnaire items, in the same format as the intrinsic motivation items, were averaged to yield a score indicative of the degree employees felt personally involved in their work. The items were selected from those used by Lodahl and Kejner (1965). Internal consistency reliability of the scale was .81. The three items are: (*a*) The most important things that happen to me involve my work. (*b*) I live, eat, and breathe my job. (*c*) I am very much personally involved in my work.

3. General job satisfaction: employees completed three questionnaire items designed to tap overall job satisfaction. The three items also were in the same format as the intrinsic motivation items and were averaged to provide an index of general satisfaction. Internal consistency reliability of the 3-item scale was .76. The items are: (*a*) Generally speaking, I am very satisfied with my job. (*b*) I frequently think of quitting my job (reversed scoring). (*c*) Generally speaking, I am very satisfied with the kind of work I have to do on my job.

The nine items used in the general satisfaction, job involvement, and intrinsic motivation scales were randomly spread throughout a 23-item section of the employee questionnaire which addressed various types of reactions to work. Lawler and Hall (1970) have argued that it is useful, on a conceptual level, to treat satisfaction, involvement, and intrinsic motivation separately. In addition, they demonstrated that the three variables do have discriminant validity in predicting other employee attitudes and behaviors—even though the variables are moderately positively intercorrelated. The correlations obtained in the present research among intrinsic motivation, general satisfaction, and the three items reflecting the specific focus of employee motivation are included in Table 5.

4. Specific satisfaction items: Employees indicated on 12 questionnaire items the degree to which they were satisfied with particular aspects of their job. The items ranged from "extremely dissatisfied" (scored 1) to "extremely satisfied" (scored 7). The 12 items were not summed but were retained as separate indicators of specific satisfactions of the job. Median intercorrelation among the 12 items was .40. The items queried how satisfied employees were with:

(*a*) The feeling of self-esteem or self-respect a person gets from being in my job.
(*b*) The opportunity for personal growth and development in my job.
(*c*) The prestige of my job inside the company (that is, the regard received from others in the company).
(*d*) The amount of close supervision I receive.
(*e*) The opportunity for independent thought and action in my job.
(*f*) The feeling of security in my job.
(*g*) The pay for my job.
(*h*) The feeling of worthwhile accomplishment in my job.
(*i*) The opportunity, in my job, for participation in the determination of methods, procedures, and goals.
(*j*) The opportunity to develop close friendships in my job.
(*k*) The opportunity for promotion.
(*l*) The amount of respect and fair treatment I receive from my boss.

Performance measures. Supervisors rated the performance of each employee on several dimensions, three of which are used in this report: (*a*) quantity of work produced, (*b*) quality of work produced, and (*c*) overall performance effectiveness.

The quantity and quality scores for each employee were obtained by averaging the ratings he received on 7-point rating scales from his supervisors (when more than one supervisory rating was available). The overall performance measure was obtained by summing average supervisory ratings across seven separate performance scales, including quantity and quality. Interjudge reliabilities of the quantity and quality scales were .71 and .79, respectively, and the internal consistency reliability of the overall performance effectiveness scale was .92. The correlation between the quantity and the quality scales was .63.

Anchors for the performance ratings were phrased in absolute terms, and supervisors were encouraged to assess each employee against the highest conceivable level of performance possible for the job in making his assessments. Nevertheless, the degree to which ratings made by different supervisors on different jobs are directly comparable is open to question, and results involving the performance ratings must be interpreted with caution.

Absenteeism measure. The number of occasions an employee was absent during the 12-mo. period during which the study took place was derived from company payroll records. Occasions absent rather than days absent were used to discount the effects of single long periods of absence. For example, if an employee were ill for a 2-wk. period, the data would show the event as one occasion of absence, rather than as 10 working days of absence.

Summary of measures. The various measures used in the research are summarized below. The means, standard deviations, and intercorrelations of all dependent variable measures are presented in Table 5. The measures include: (*a*) Job descriptions: variety, autonomy, task identity, feedback, dealing with others, and friendship opportunities. Descriptions were made by employees, by supervisors, by the researchers using the Turner-Lawrence procedures, and by the researchers subjectively after job observations and interviews. Because there was generally high convergent validity among the four sets of job descriptions, and because the conceptual basis of the study suggests that jobs as experienced by employees should be most directly causal of employee reactions to their jobs, the employee job descriptions are used in the analyses to be reported in subsequent sections. (*b*) Level of higher order need strength of employees: a summary score of employee reactions to 12 need strength items. (*c*) Employee reactions to their jobs and work. Questionnaire-derived measures of: the amount of intrinsic motivation experienced by employees, and the focus of their motivation; general job satisfaction; personal job involvement; and 12 specific satisfactions or dissatisfactions with the job or the work situation. (*d*) Rated performance effectiveness of employees, in terms of quantity of work produced, work quality, and overall performance effectiveness. (*e*) Absenteeism, measured by the number of occasions an employee was absent during a 12-mo. period.

RESULTS

General Relationships between Job Characteristics and Employee Reactions

According to the conceptual position outlined earlier, the nature of the relationships between job characteristics and employee reactions to their work (including satisfaction, performance, and absenteeism) will depend upon the need states of the employees. In particular, it was predicted that, if employees are desirous of higher order need satisfactions, there should be a positive relationship between the four core dimensions (variety, autonomy, task identity, and feedback) and motivation, satisfaction, performance, and attendance.

The mean score of the 208 telephone company employees on the 12-item scale indicative of higher order need strength was 6.01. Given that the maximum possible score on the scale is 7.0, it would appear that the employees who participated in the present research had strong higher order needs, or at least that they felt

TABLE 5

Intercorrelations among Dependent Variables

Variable	$\bar{X}$	SD	Correlations																					
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. Level of intrinsic motivation	5.96	.87	—																					
Focus of motivation																								
2. Taking personal responsibility	6.39	.99	.30*	—																				
3. Doing large quantities of work	5.21	1.53	.01	.17*	—																			
4. Doing high quality work	5.22	1.09	.25*	.44*	.08	—																		
Rated performance																								
5. Quantity	5.24	1.45	.04	−.09	.09	−.05	—																	
6. Quality	5.24	1.27	.13	.05	.08	−.04	.63*	—																
7. Overall effectiveness	5.53	1.11	.18*	.02	.10	.06	.79*	.85*	—															
8. General job satisfaction	4.97	1.26	.39*	.18*	.01	.23*	.07	.08	.16*	—														
9. Job involvement	3.00	1.37	.39*	.20*	.04	.16*	.07	.07	.11	.44*	—													
10. Absenteeism (no. of times absent)	2.69	2.16	−.23*	−.14*	−.10	−.06	−.12	−.26*	−.33*	−.10	−.15*	—												
Specific satisfaction items																								
11. Self-esteem obtained from job	4.34	1.44	.23*	.04	−.01	.06	−.07	−.02	−.04	.49*	.34*	−.06	—											
12. Personal growth and development	4.33	1.48	.26*	.18*	.04	.11	.07	.03	.07	.58*	.38*	−.02	.66*	—										
13. Prestige of job inside company	3.95	1.44	.16*	−.01	.10	.04	−.09	−.03	−.03	.44*	.29*	−.09	.63*	.50*	—									
14. Amount of close supervision received	4.46	1.45	.25*	.14*	−.01	.13	.04	.06	.10	.28*	.25*	−.07	.31*	.45*	.30*	—								
15. Independent thought and action	4.62	1.58	.21*	.13	−.11	.10	.11	.12	.17*	.36*	.34*	−.09	.39*	.50*	.32*	.62*	—							
16. Security	5.32	1.47	.15*	.13	.02	.16*	−.01	.06	.09	.47*	.23*	−.08	.43*	.46*	.39*	.40*	.40*	—						
17. Pay	3.68	1.74	.07	.04	−.02	.10	−.06	−.13	−.11	.44*	.16*	.02	.41*	.43*	.39*	.21*	.15*	.45*	—					
18. Feeling of worthwhile accomplishment	4.82	1.61	.45*	.22*	.01	.19*	.09	.14*	.11	.58*	.39*	−.02	.58*	.58*	.41*	.32*	.49*	.53*	.42*	—				
19. Participation in job-related decisions	3.86	1.50	.23*	.08	.05	.13	.04	.04	.07	.31*	.28*	−.04	.38*	.52*	.39*	.37*	.46*	.30*	.31*	.50*	—			
20. Development of close friendships	4.78	1.28	.06	.06	−.11	−.02	−.09	−.04	−.14*	.19*	.05	.15*	.24*	.34*	.21*	.25*	.31*	.28*	.22*	.33*	.33*	—		
21. Promotion	3.98	1.64	.16*	.16*	.08	.15*	.07	−.04	.03	.39*	.34*	.09	.43*	.60*	.42*	.38*	.29*	.44*	.48*	.43*	.51*	.22*	—	
22. Respect and fair treatment from boss	5.34	1.47	.23*	.08	−.06	.14*	.10	.06	.15*	.33*	.10	−.07	.42*	.40*	.35*	.43*	.41*	.57*	.34*	.45*	.35*	.25*	.42*	—

Note.—N = 208.
* $p < .05$ (two-tailed test).

TABLE 6

GENERAL RELATIONSHIPS BETWEEN JOB CHARACTERISTICS AND EMPLOYEE REACTIONS

Dependent variable	Core dimension				Interpersonal dimension	
	Variety	Autonomy	Task identity	Feedback	Dealing with others	Friendship opportunities
Level of intrinsic motivation	.32*	.30*	.16*	.18*	.07	.09
Focus of motivation						
Taking personal responsibility	.14*	.12*	.19*	.06	.08	.05
Doing large quantities of work	−.10	−.12	.01	.02	.06	−.17
Doing high quality work	.16*	.12*	.13*	.10	.04	.12*
Rated performance						
Quantity	−.03	.13*	.05	.00	−.02	−.18
Quality	.17*	.16*	.07	.02	−.11	−.02
Overall effectiveness	.20*	.26*	.11*	−.03	−.07	−.09
General job satisfaction	.38*	.39*	.20*	.28*	.17*	.21*
Job involvement	.24*	.22*	.12*	.24*	.03	.16*
Absenteeism (no. of times absent)	.02	−.14*	−.22*	−.10	.01	−.05
Specific satisfaction items						
Self-esteem obtained from job	.32*	.32*	.15*	.35*	.15*	.27*
Personal growth and development	.36*	.34*	.14*	.31*	.11*	.29*
Prestige of job inside company	.30*	.25*	.15*	.35*	.14*	.28*
Amount of close supervision received	.31*	.35*	.13*	.30*	.07	.16*
Independent thought and action	.53*	.62*	.25*	.15*	.00	.25*
Security	.22*	.27*	.19*	.39*	.15*	.28*
Pay	.04	.05	.04	.34*	.25*	.24*
Feeling of worthwhile accomplishment	.29*	.32*	.28*	.42*	.23*	.31*
Participation in job-related decisions	.28*	.27*	.20*	.34*	.12*	.25*
Development of close friendships	.25*	.12*	.09	.29*	.09	.47*
Promotion	.17*	.20*	.15*	.34*	.21*	.19*
Respect and fair treatment from boss	.19*	.26*	.22*	.35*	.14*	.24*

Note.—$N = 208$.
* $p < .05$ (one-tailed test).

it was appropriate to express high desires for these need satisfactions. Therefore, generally positive relationships would be expected between job descriptions on the four core dimensions and the dependent variables.

Correlations between the six job description measures obtained from employees and the scores of employees on the dependent variable measures are presented in Table 6. Results are discussed below separately for the four core dimensions and the two interpersonal dimensions.

Core Dimensions: Toward Internal Motivation

In general, positive relationships were obtained between the four core dimensions and dependent measures indicative of motivation, satisfaction, performance, and attendance.

Performance motivation and actual performance. The higher jobs are on the core dimensions, the more employees tend to report feeling internal pressures to take personal responsibility for their work and to do high quality work. And, in fact, when jobs are described as being higher on variety, autonomy, and task identity, employees are rated as doing higher quality work and as being generally more effective performers on the job.

Further, the data provide support for the interpretation of these relationships which was proposed earlier in this paper: when jobs are high on the core dimensions, employees report having higher intrinsic motivation to perform well. That is, employees indicate that when they perform well on the job, they experience positive internal feelings, and when they perform poorly, they feel badly. Apparently jobs seen by employees as being high on the core dimensions do establish conditions on the job so that it is possible for workers to obtain personally rewarding experiences by doing well on the job. The data suggest, moreover, that "doing well" is interpreted in the job context as having much more to do with high quality performance than producing large quantities of work. The core dimensions do not relate either to felt pressure for high quantity production, or to the actual quantity of work which is produced. This is consistent with the notion that employees with strong higher order needs feel positively when they have "accomplished something that they feel is meaningful"; it is not unreasonable that such workers would see doing high quality work as much more of a meaningful "accomplishment" than simply turning out large quantities of work.

General satisfaction, job involvement, and absenteeism. The core dimensions are, as expected, strongly and positively related to overall job satisfaction and to the degree that employees feel personally involved in their work. It is not, therefore, surprising to find that employees whose jobs are seen as high on the core dimensions tend to have better absence records as well. Work apparently is a satisfying place to be for employees with jobs high on the four core dimensions, and one way they behaviorally demonstrate this is by coming to work regularly.

Specific satisfactions. Nearly all of the specific satisfaction items were significantly positively correlated with the descriptions of jobs on the core dimensions. This is to be expected from the fact that the core dimensions relate substantially to overall job satisfaction—since overall satisfaction is likely to be strongly influenced by satisfaction with the particular aspects of the work situation addressed by the 12 specific satisfaction items.

It is instructive, however, to note which of the specific satisfaction items are especially strongly related to the core dimensions, and which have negligible relationships. Items with consistently strong relationships to the core dimensions should be considered especially sensitive to differences on these dimensions, whereas those items with weak relationships should be considered not as responsive to differences in job design. These data should be helpful, therefore, in more fully understanding the kinds of effects that the core dimensions have on employees' reactions to their jobs.

The four specific satisfaction items most strongly related to the core dimensions are (in descending order): (*a*) The opportunity for independent thought and action in my job. (*b*) The feeling of worthwhile accomplishment in my job. (*c*) The opportunity for personal growth and development in my job. (*d*) The self-esteem and self-respect a person gets from being in my job.

The four items least strongly related to the core dimensions are (in ascending order): (*a*) The pay for my job. (*b*) The opportunity to develop close friendships in my job. (*c*) The opportunity for promotion. (*d*) The amount of respect and fair treatment I receive from my boss.

It appears, as expected, that the four core dimensions seem to be most strongly related to the satisfaction of higher order needs. The satisfaction items which are most weakly related to the four core dimensions seem to be most relevant to the satisfaction of needs classified as lower-order in the hierarchies of Maslow (1943) and Alderfer (1969).

Two Interpersonal Dimensions

Also included in Table 6 are the relationships between two primarily interpersonal job dimensions (dealing with others, and friendship opportunities) and the dependent variable measures. The nature of the relationships obtained are strikingly different from those obtained for the four core dimensions. The two interpersonal dimensions do not relate very consistently or strongly either to employee affective responses to the job, or to their actual work performance. There is no significant relationship between the dimension "dealing with others" and any measure of motivation, performance, or absenteeism.

Relationships involving the dimension "friendship opportunities" also are generally negligible. The dimension does not relate to the level of intrinsic motivation employees report, or to any measure of performance or absenteeism.

There are positive relationships (moderately large for friendship opportunities, small for dealing with others) between the two interpersonal dimensions and job satisfaction. The single largest relationship between any of the satisfaction items and either of the interpersonal job dimensions is not a very surprising one: The more friendship opportunities are present on a job, the more employees are satisfied with the chances they have to develop close friendships on the job. This relationship suggests that the kinds of consequences to be expected from having jobs with high interpersonal components (as measured in this research) are primarily social in nature—rather than being relevant to the performance and motivation of employees as is the case for jobs high on the core dimensions.

Within-Job Relationships between Job Dimensions and Employee Reactions

It was proposed earlier that job design factors affect employee attitudes and behavior because of their impact on the perceptions employees have of their jobs. This may be represented schematically as follows:

$$\begin{array}{c}\text{Objective job}\\ \text{characteristics}\end{array} \xrightarrow{1} \begin{array}{c}\text{Perceived job}\\ \text{characteristics}\end{array} \xrightarrow{2} \begin{array}{c}\text{Work attitudes}\\ \text{and behavior}\end{array}$$

Data presented in Table 2 show that employee perceptions of their jobs have substantial convergence with the assessments of objective job characteristics made by the researchers and by company supervisors. If it is assumed that the assessments by the researchers and supervisors do reflect the "objective" character of the jobs, these data strongly suggest that employee job perceptions are based (at least in major part) on objective job characteristics—the first link in the chain. Results reported in Table 6 and discussed immediately above show that employee job perceptions also are related to work attitudes and behavior—the second link in the chain. However, these latter analyses are based on job descriptions provided by employees who worked on 13 different jobs—jobs which have been shown (see Table 4) to vary substantially on the six job dimensions. These large objective differences among the 13 jobs may, therefore, have been primarily responsible for the relationships which were obtained. In other words, the results in Table 6 could have come about because the objective job characteristics directly caused both the perceived job characteristics and the employee reactions—and not because they caused the perceived job characteristics which, in turn, caused the reactions.

By examining the relationships between employee perceptions of their jobs and the dependent variables separately for workers on each of the 13 jobs, it is possible to determine if perceptual factors, largely uncontaminated by objective between-job differences, bear a significant relationship to the dependent variables. Such an analysis, therefore, would provide a relatively unconfounded test of the second link in the above diagram. Within-job relationships should, of course, be lower in magnitude than those obtained using all employees working on all jobs, since there is little meaningful variation in actual job characteristics when data are obtained only from a single job, and as a result the variance in employee perceptions should be severely attenuated.[8]

Thirteen matrices of relationships were obtained by correlating job descriptions and dependent variables separately for employees working on each of the 13 jobs. The median of the 13 correlations between each job dimension and each dependent variable was selected, and the matrix of median correlations is presented in Table 7.

As expected, the median within-job correlations generally are consistent with the correlations computed across all employees and jobs (see Table 6). The order of magnitude of the median within-job correlations, also as expected, is lower than that of the correlations based on data from all employees. These results suggest, therefore, that employees' perceptions of their jobs are of central importance in affecting job attitudes and behaviors, but that the major determinant of such perceptions is the objective make-up of the job itself.

---

[8] To the extent that there are differences in the actual jobs assigned to employees who work on the same objective job, the usefulness of within-job analyses for testing the impact of strictly perceptual factors on the dependent variables is limited. Although some differences in actual work assignments were noted by the researchers among employees working on the same objective job, these differences invariably were of little consequence. Therefore, the within-job analyses probably do reflect primarily differences in employee job perceptions rather than actual differences in job design.

TABLE 7

MEDIAN WITHIN-JOB CORRELATIONS BETWEEN JOB CHARACTERISTICS AND EMPLOYEE REACTIONS

Dependent variable	Core dimension				Interpersonal dimension	
	Variety	Autonomy	Task identity	Feedback	Dealing with others	Friendship opportunities
Level of intrinsic motivation	.31	.23	.11	.35	.03	.06
Focus of motivation						
Taking personal responsibility	.00	.00	.18	.09	.15	.00
Doing large quantities of work	−.05	−.06	.10	.01	.07	−.07
Doing high quality work	.05	.22	.19	.05	.10	.06
Rated performance						
Quantity	.06	.21	−.10	−.05	−.02	−.18
Quality	.06	.18	.06	.03	−.10	−.05
Overall effectiveness	.12	.21	.06	.04	.00	−.11
General job satisfaction	.38	.48	.21	.39	.22	.21
Job involvement	.34	.28	.20	.19	.22	.04
Absenteeism (no. of times absent)	−.10	−.03	.03	.00	.17	.14
Specific satisfaction items						
Self-esteem obtained from job	.26	.31	.15	.44	.18	.27
Personal growth and development	.57	.40	.18	.34	.25	.21
Prestige of job inside company	.44	.44	.17	.42	.24	.26
Amount of close supervision received	.19	.39	.08	.31	.17	.07
Independent thought and action	.40	.55	.31	.30	.13	.20
Security	.21	.28	.28	.35	.11	.25
Pay	.24	.33	.17	.33	.19	.18
Feeling of worthwhile accomplishment	.40	.37	.41	.48	.25	.26
Participation in job-related decisions	.34	.34	.39	.31	.31	.25
Development of close friendships	.27	.07	.25	.35	.26	.51
Promotion	.19	.33	.13	.32	.26	.06
Respect and fair treatment from boss	.20	.29	.24	.35	.09	.30

Note.—Because the correlations reported in this table are medians, no indication of statistical significance is given.

Exploration of Condition for Enhancing Motivation and Satisfaction

Results presented in the preceding discussion indicate that when jobs are high on the core dimensions, employees who have strong desires for higher order need satisfaction will be highly motivated and well satisfied on the job. In the following discussion, two analyses will be reported which extend these findings and provide additional documentation of the conditions which seem necessary to enhance employee motivation and satisfaction through job design.

Must Jobs Be High on All Four Core Dimensions?

The theory on which the present research is based specifies that when jobs are high on the core dimensions, employees have the opportunity to find out (feedback) that they personally (autonomy) have accomplished something meaningful (task identity and variety) when they perform well. The implication of this assertation is that, for maximum motivation, jobs should be high simultaneously on all four of the core dimensions

To test this possibility, *S*s were partitioned into three groups: (*a*) those who described their jobs as being above the 60th percentile on all four core dimensions; (*b*) those who described their jobs as being below the 40th percentile on all four core dimensions; and (*c*) the majority of *S*s, who typically described their jobs as being high on some of the core dimensions and low on others. A one-way, missing-data, unequal *N* analysis of variance was used to compare the mean scores of em-

TABLE 8

COMPARISON OF JOBS DESCRIBED AS HIGH ON ALL CORE DIMENSIONS, HIGH ON SOME CORE DIMENSIONS, AND LOW ON ALL CORE DIMENSIONS

Dependent variable		Dependent variable $\bar{X}$		
	F ratio	Jobs low on all core dimensions[a]	Jobs high on some core dimensions[b]	Jobs high on all core dimensions[c]
Level of intrinsic motivation	13.72**	5.28	5.99	6.46
Focus of motivation				
Taking personal responsibility	6.08**	5.90	6.40	6.88
Doing large quantities of work	.24	5.48	5.17	5.29
Doing high quality work	3.04*	6.00	6.20	6.71
Rated performance				
Quantity	.15	5.19	5.23	5.39
Quality	3.36*	4.97	5.57	5.89
Overall effectiveness	3.40*	5.14	5.54	5.94
General job satisfaction	28.97**	3.78	4.99	6.21
Job involvement	18.40**	2.30	2.94	4.43
Absenteeism (no. of times absent)	1.62	3.11	2.70	2.00
Specific satisfaction items				
Self-esteem obtained from job	27.06**	3.33	4.30	6.00
Personal growth and development	24.45**	3.43	4.27	6.00
Prestige of job inside company	21.65**	3.24	3.88	5.59
Amount of close supervision received	15.97**	3.57	4.44	5.65
Independent thought and action	31.66**	3.24	4.62	6.35
Security	9.75**	4.52	5.32	6.24
Pay	2.95	3.38	3.63	4.47
Feeling of worthwhile accomplishment	35.59**	3.05	4.88	6.41
Participation in job-related decisions	21.09**	2.81	3.85	5.29
Development of close friendships	9.56**	4.14	4.77	5.65
Promotion	13.98**	3.24	3.92	5.47
Respect and fair treatment from boss	14.64**	4.14	5.40	6.24

[a] $N = 21$.
[b] $N = 170$.
[c] $N = 17$.
* $p < .05$ (one-tailed test).
** $p < .01$ (one-tailed test).

ployees in the three groups on each of the dependent variables. Means and *F* ratios are presented in Table 8.

Results are consistent with those reported in Table 6 and, in addition, tend to be more substantial in magnitude and more statistically reliable.[9] It was expected that when jobs were described as high on all four core dimensions: (*a*) experienced pressures to take personal responsibility for one's work and to do high quality work would be high, (*b*) intrinsic motivation would be high, (*c*) rated performance quality and overall performance effectiveness would be high, (*d*) job satisfaction and involvement would be high, and (*e*) absenteeism would be low. All of these expectations were borne out substantially and statistically significantly, except for the one involving absenteeism. While absenteeism was lowest when jobs were seen as high on all four core dimensions, differences among the three groups were not statistically significant.

Although these results support the contention that motivation, satisfaction, and performance are enhanced when jobs are high simultaneously on all four core dimensions, they do not imply that a disjunctive model for combining the core dimensions is necessarily the optimal one. Therefore, the relative usefulness of other ways of combining the dimensions was explored. In particular, (*a*) an unweighted sum of the four dimensions was computed for each *S* and correlated with each of the dependent variables, (*b*) the four dimensions were used as independent variables in a multiple-regression analysis predicting each of the dependent variables, and (*c*) the product of the four dimensions (Variety × Autonomy × Task Identity × Feedback) was correlated with each of the dependent variables. The latter score is a continuous measure which is congruent with the disjunctive combinatory model specified by the theory. That is, for a product score to be high, all of the four core dimensions must be moderately high, and, therefore, the measure reflects the degree to which a job is high simultaneously on all four core dimensions. Correlations involving the product score were compared to those obtained using the other two combinatory models.

---

[9] The analytic procedures used, unfortunately, do not permit computation of the amount of variation controlled, for comparison with the correlational data reported earlier.

Correlations between the unweighted sum score and the dependent variables ranged from −.08 to .61, with a median of .30. Multiple correlations ranged from .13 to .65, with a median of .37. Correlations using the product score ranged from −.05 to .60, and also had a median of .30. Since some shrinkage would be expected in the magnitude of the multiple correlations obtained upon cross-validation, it seems appropriate to conclude that the results did not meaningfully differentiate among the three means of combining the core dimensions. Thus, while the theory-specified disjunctive model is not disconfirmed by the data, neither has it been shown to represent a more adequate means of combining the dimensions than the other models tested.

Does High Order Need Strength Moderate Obtained Relationships?

It has been maintained throughout this article that jobs which are high on the core dimensions should be motivating only to individuals who are desirous of the intrinsic rewards that the jobs provide, namely, higher order need satisfactions.

The analysis reported in this section tests the possibility that *S*s differing in higher order need strength do in fact show differential responsiveness to jobs high on the core dimensions, despite the generally high mean of all *S*s on the need strength measure. Correlations between each of the core dimensions and the dependent variables were computed separately for those employees whose higher order need strength scores were in the top one-third of the distribution of scores for all *S*s, and for those employees whose scores were in the bottom one-third of the same distribution. The mean higher order need strength score for the employees in the top third was 6.78, and the mean score for employees in the bottom third was 5.09. Since 5.09 is still one full point above the midpoint of the scale, it was expected that even for these "low" need strength employees, relationships between the core dimensions and the dependent variables would be in the positive direction, but lower in magnitude than the relationships for *S*s in the top third of the distribution. It also was expected that the relationship found for the top third should be stronger than those reported in Table 6 for all *S*s. Results are presented in Table 9.

For variety, autonomy, and, to a lesser extent, feedback, the expectation of a differential relationship was confirmed. Relationships between these job dimensions and the dependent variables are larger, and often strikingly so, for the *S*s with high higher order need strength than for *S*s with moderate higher order need strength. Further, the correlations for higher need strength employees are typically larger than those reported in the general analysis involving all *S*s (see Table 6), and the correlations based on the lower need strength *S*s are typically lower than the general relationships.

The moderating effect of higher order need strength was not obtained for task identity. Part of the reason for this may have to do with the nature of the particular jobs included in this study. Among the jobs in this study which were high on task identity were those of telephone operators. By and large, when an operator takes a toll call she does a "whole" piece of work. She is involved in the call from the time the customer comes on the line until the called party answers and confirms his identity. The operator will do and redo this task and similar tasks many times during a typical working day. Observations by the researchers strongly suggest that an operator with strong higher order need strength quickly becomes bored and impatient with so many repetitions of this small but "whole" piece of work. If this is so, then it should not be sur-

TABLE 9

MODERATING EFFECTS OF HIGHER ORDER NEED STRENGTH

Dependent variable	Variety		Autonomy		Task identity		Feedback	
	High strength	Low strength	High strength	Low strength	High strength	Low strength	High strength	Low strength
Level of intrinsic motivation	.56*	.15	.49*	.07	.19	.22*	.41*	.25*
Focus of motivation								
Taking personal responsibility	.41*	.00	.39*	.07	.30*	.23*	.21*	−.02
Doing large quantities of work	−.07	−.20	−.19	−.06	−.18	.03	.10	.09
Doing high quality work	.17	.19	.18	.03	.07	.01	.21*	.09
Rated performance								
Quantity	−.02	−.24	.09	−.09	.01	.15	.04	.23*
Quality	.15	.05	.17	−.05	−.02	.31*	.12	.03
Overall effectiveness	.09	.05	.16	−.01	.01	.26*	−.04	.06
General job satisfaction	.41*	.28*	.43*	.29*	.27*	.18	.17	.33*
Job involvement	.30*	.17	.35*	.19	.16	.26*	.30*	.15
Absenteeism (no. of times absent)	−.29*	−.18	−.24*	.01	.21	−.26*	.04	.06
Specific satisfaction items								
Self-esteem obtained from job	.48*	.06	.39*	.07	.38*	.12	.34*	.29*
Personal growth and development	.55*	.03	.47*	−.01	.34*	.04	.39*	.33*
Prestige of job inside company	.41*	.17	.32*	.07	.32*	.18	.41*	.24*
Amount of close supervision received	.33*	.15	.43*	.10	.06	.07	.45*	.28*
Independent thought and action	.62*	.42*	.75*	.45*	.28*	.29*	.37*	−.01
Security	.29*	.04	.38*	.08	.20*	.24*	.41*	.33*
Pay	.24*	−.10	.15	.09	.07	.03	.26*	.36*
Feeling of worthwhile accomplishment	.53*	.06	.49*	.15	.38*	.33*	.42*	.40*
Participation in job-related decisions	.41*	.20*	.32*	.18	.15	.22*	.41*	.25*
Development of close friendships	.56*	−.04	.40*	−.35*	.38*	−.03	.36*	.13
Promotion	.29*	−.09	.35*	−.10	.14	.10	.40*	.34*
Respect and fair treatment from boss	.39*	.06	.45*	.03	.21*	.22*	.38*	.38*

Note.—$N = 67$ in each group.
* $p < .05$ (one-tailed test).

prising to find that there is little difference in the relationship between task identity and the dependent variables for employees with high versus low higher order need strength.

Indeed, it would be expected from the theory that differences between high and low need strength employees should be maximized when jobs are moderately high on all four core dimensions. The job of the toll operator clearly does not meet this criterion: while it is reasonably high on task identity and feedback, it is very low on both variety and autonomy (see Table 4).

To examine the validity of this argument, therefore, the product score (Variety × Autonomy × Task Identity × Feedback) described earlier was correlated with the dependent variables separately for the third of *S*s highest in higher order need strength and for the third of *S*s lowest in higher order need strength. Since the product score reflects the degree to which a job is seen as being simultaneously high on all four core dimensions, substantial differences would be expected in the magnitude of correlations obtained for high versus low need strength employees. Results are presented in Table 10.

Again, obtained relationships were much higher for *S*s who were high in relevant need strength than for *S*s low in need strength for nearly all of the measures of motivation, satisfaction, and performance. The difference be-

TABLE 10

MODERATING EFFECT OF HIGHER ORDER NEED STRENGTH FOR JOBS HIGH ON ALL FOUR CORE DIMENSIONS

Dependent variable	High need strength	Low need strength
Level of intrinsic motivation	.54*	.23*
Focus of motivation		
Taking personal responsibility	.37*	.12
Doing large quantities of work	−.09	−.01
Doing high quality work	.21*	.12
Rated performance		
Quantity	.07	−.04
Quality	.23*	.02
Overall effectiveness	.15	.05
General job satisfaction	.48*	.40*
Job involvement	.45*	.28*
Absenteeism (no. of times absent)	−.26*	−.08
Specific satisfaction items		
Self-esteem obtained from job	.54*	.13[a]
Personal growth and development	.57*	.16[a]
Prestige of job inside company	.50*	.26*
Amount of close supervision received	.48*	.27*
Independent thought and action	.70*	.45*[a]
Security	.49*	.25*[a]
Pay	.27*	.09
Feeling of worthwhile accomplishment	.59*	.32*[a]
Participation in job-related decisions	.44*	.37*
Development of close friendships	.54*	−.19[a]
Promotion	.45*	.08[a]
Respect and fair treatment from boss	.48*	.20*[a]

Note.—$N = 67$ in each group.
[a] The difference between the correlations for high need strength *S*s and for low need strength *S*s is significant at $p < .05$.
* $p < .05$ (one-tailed test).

tween the correlations for high and low need strength *S*s was tested for statistical significance, and these results are included in the table. Even though most of the correlations were positive (because of the relatively high mean need strength for even the "low" need strength employees), a number of pairs of correlations were significantly different from one another. All in all, the data make a strong case for the moderating effect of individual higher order need strength in determining the effects of job characteristics on employee behavior and attitudes at work.

DISCUSSION AND IMPLICATIONS

Job Design and Individual Differences

The results of this study suggest that there are important interdependencies among the characteristics of individuals and the characteristics of jobs which must be taken account of in the development of any full understanding of the impact of various kinds of job designs. Both the advocates of a "scientific management" approach to job design (make the work routine, simple, and standardized) and the more recent supporters of "job enlargement" (make the work complex, challenging, and demanding of individual responsibility and decision making) appear to have attached insufficient importance to individual-job interactions in determining affective and behavioral reactions to jobs. Those of the scientific management persuasion, for example, have tended to assume that the typical employee will be content, if paid judiciously for his cooperation, to work on jobs which provide little or no opportunity for personal feelings of accomplishment or achievement. Those of the job enlargement school, on the other hand, have tended to assume that most employees are desirous of such opportunities and will work hard and effectively when they have a job which provides them. The present research suggests that, depending on the characteristics of the workers involved, both points of view would lead to job design practices which are appropriate some of the time and inappropriate other times.

The present findings and conclusions fit well with the previous research of Turner and Lawrence (1965) and Hulin and Blood (1968). In both of these studies a sociological-level variable (urban vs. rural background) was shown to moderate the relationship between job characteristics and employee satisfaction. The relationship between job level (i.e., jobs which would be high on the four core dimensions) and satisfaction was found to be high for rural workers and low for urban workers. The present study demonstrates that individual higher order need strength also moderates the relationship between job level and satisfaction (as well as the relationship between job level and other dependent variables). In particular, individuals with strong desires for higher order need satisfaction respond much more positively to high level jobs than do individuals who have weaker higher order needs.

It may be that urban workers desire higher order need satisfaction less than do rural workers—in which case the findings of the earlier studies and the present study would be highly congruent. This possibility was tested with the present data and a small difference in the expected direction was in fact found: workers with rural backgrounds were higher on higher order need strength than were workers with urban backgrounds ($t = 1.47$, $p < .10$). Thus, the present research has taken the previous finding regarding urban–rural differences from the sociological level to the psychological level, refined it, and successfully replicated it. The present results also substantially extend the previous results in that the moderating effect of individual differences has been shown to apply to a number of dependent variables in addition to job satisfaction (i.e., job involvement, work motivation, quality of performance). Presumably it now should be possible to make considerably more specific predictions about how different individuals will respond to jobs with various characteristics.

It has been assumed throughout the above discussion that job characteristics actually cause the differences in employee satisfaction, motivation, performance, and absenteeism which were observed. Although the predictions which were made (and confirmed by the data) were based on a conceptual framework which includes causal propositions, the study design was correlational and at no point were the causal links in the theory directly tested. There is, however, reason to believe that the causal directions specified in the theory may be correct. While the simple correlations between job characteristics and the dependent variables (see Table 6) are open to alternative causal interpretations, this appears to be much less the case for those data which support other, more specific, predictions of the theory. In particular, it is difficult even after the fact to imagine an alternative causal model which would predict simultaneously that (*a*) all four of the core dimensions need to be present to realize the most positive affective and behavioral reactions on the part of employees, and that (*b*) the strength of desire for higher order need satisfaction will substantially moderate the relationships between the core dimensions and the dependent variables. It appears, therefore, that a cautious interpretation of the results in terms of the causal impact of job characteristics may be reasonable. This does not deny the fact that any theory can best be tested by experimental alteration of the independent variables, and manipulation of the core dimensions in an experimental setting would seem to be a clear next step in research on job effects.

Use of Perceived Job Characteristics

One of the major conceptual and methodological problems which pervades studies of task and job effects on behavior has to do with the differences between task materials as they exist in objective reality and as they are perceived by individual performers. Tasks and jobs are invariably redefined by the individuals who perform them, sometimes deliberately and sometimes without full awareness by the performers of the changes or re-emphases that are being made. Further, it is the redefined task rather than the objective task which the individual tries to perform, and thus only those aspects of tasks or jobs which are actually perceived or experienced by a performer can have an impact on his performance and attitudes. This would seem to argue for the use of task and job characteristics as described by the performers themselves in research aimed at ascertaining the effects of these characteristics on performance. Yet when such subjective assessments are used, many of the important conceptual and methodological advantages associated with the use of independently and objectively described independent variables are lost.[10]

There are at least two strategies for dealing with this problem in research on task and job effects on behavior. One, which was used in the present research, is to employ subjective assessments of the tasks or jobs by the performers themselves, but simultaneously to develop means of determining the relationship between these assessments and others, including objective measures when possible. Thus, in addition to employee descriptions of the characteristics of their jobs in the present study, jobs were rated by three other methods: by company managers, by the researchers, and by an adaptation of the Turner and Lawrence (1965) procedures for describing jobs in operational terms. As was shown in Table 2, there was high convergence among the four methods for all job dimensions except feedback. Thus, while the job descriptions used in the analyses are those actually experienced by the employees (and thus presumably are more directly causal of their behavior than would be the case for objectively described characteristics), they also have been shown to have substantial convergent validity and to be based upon the objective job characteristics.

A second strategy for dealing with the redefinition problem has been proposed by Hackman (1970) and, while methodologically more difficult to implement, probably has more potential for the development of a general understanding of the effects of tasks and jobs. In essence, this strategy suggests that the redefinition process should be viewed as the first stage of the performance process itself, and the redefined task should be conceived of as a potentially measurable intervening variable in the causal chain between the objective task input and the dependent variables of interest (e.g., performance, satisfaction). Just as individual differences in need strength have been shown in this research to moderate the effects of job characteristics on employee behavior and attitudes, so would individual needs, values, and goals be expected to interact with the objective task or job in influencing task redefinition. It might be, for example, that individuals with a particular pattern of needs tend to redefine the tasks they perform to be more consistent with those needs than is actually the case. These individuals might then develop hypotheses about appropriate on-the-job behavior which they expect to lead to need-satisfying outcomes, but which, because of the lack of congruence between the objective and the redefined task, might in fact serve only to set the stage for disillusionment and frustration. In any case, it appears that research on the ways different individuals redefine tasks and jobs may have considerable potential for furthering general understanding of the effects of tasks and jobs on the behavior and attitudes of performers.

---

[10] The conceptual and methodological implications of the task redefinition phenomenon are discussed in more detail elsewhere (Hackman, 1969b; Weick, 1965).

Nature of the Four Core Dimensions

The results of the present research show that, in general, employees with moderately high desires for higher order need satisfaction tend to work harder and be more satisfied when they perceive their jobs as being relatively high on the four core dimensions. In addition, it was shown that for the most favorable outcomes, jobs need to be at least moderately high on all four of the dimensions.

Further refinement of the core dimensions and exploration of their impact on individuals clearly is called for by the research. For example, the relationship between variety and task identity needs further examination. In one sense, since both of these dimensions are viewed as enhancing the meaningfulness of the work (although by different means), it might be argued that it is not necessary for both to be present (i.e., that one can effectively substitute for the other). The finding of Hall and Lawler (1970), that jobs which were very high on variety and low in task identity are not associated with high quality performance among researchers, would tend to argue against this possibility, however. An alternative possibility which would not be contrary to the Hall and Lawler findings is that some moderate level of variety is essential simply to keep the employee from being bored with his work (Scott, 1966), and that once this is achieved, experienced meaningfulness of the work will vary directly with the amount of task identity present. Finally, it may be that increases in variety can serve two different functions, depending on the amount initially present on a given job. For low initial levels of variety an increase may serve mainly to decrease the monotony of the work and make it possible for perceived meaningfulness to vary as a function of the amount of task identity present, as suggested above. At some point, however, variety may assume a different function, namely, introducing challenge into the work which, when successfully dealt with, can be satisfying to many individuals. Clearly, additional research on the psychological impact of these dimensions is called for.

Additional work also needs to be done on the operationalization of the feedback dimension, since no convergence was obtained among the several methods for measuring this dimension. The fact that the dimension operated as predicted in the data analyses indicates that the employee judgments of the dimension have some validity, but because the various groups who rated the jobs did not agree on the amount of feedback present, the exact meaning of the dimension remains highly uncertain.

Finally, the generality of the moderately high correlation ($r = .67$) between variety and autonomy should be tested. The descriptions of the scales used to measure these two dimensions were quite dissimilar, and it may be that the high correlation was obtained simply because jobs which had high variety in the particular organization where the study took place also tended to be jobs which provided considerable autonomy. The fact that most telephone traffic jobs (i.e., various kinds of operators) were low on both variety and autonomy would certainly facilitate obtaining a positive correlation between the dimensions, and might suggest that the high correlation is to some extent organization specific. There is reason, however, to believe that the positive relationship between variety and autonomy is at least moderately general. Turner and Lawrence (1965) found an analogously high relationship between these two dimensions across 47 different jobs, and Alderfer (1967) obtained a very high correlation between variety and "decision time" (which has some conceptual similarity to autonomy) across a sample of 30 jobs which included management jobs. It appears well worth examining the possibility that the psychological meaning of the two dimensions is similar in ways that have not been attended to in the present conceptualization.

Interpersonal Components of Jobs

The two job dimensions reflective of the interpersonal components of jobs (dealing with others, and friendship opportunities) did not relate to employee work motivation or performance. The dimensions did relate positively to certain kinds of satisfaction, but the relationships were not as substantial as those involving the four core dimensions.

According to the conceptualization on which the present study is based, the degree to which jobs require interpersonal activities should relate to work motivation only when (*a*) workers have high desires for the satisfaction of social needs, and (*b*) working hard on the job can lead to the satisfaction of these needs. Even those jobs which scored relatively high on the two interpersonal dimensions (e.g., some operators jobs) fail to meet the latter criterion. Operators reported in fact that they could obtain social satisfactions best when they were not "working hard" by company standards (i.e., completing a large number of calls), since when the load of calls was heavy they had little or no time for meaningful interpersonal activities with either customers or fellow employees.

Even if jobs were designed so that relating meaningfully to others was an integral part of doing the job well, there is reason to doubt whether such jobs would have long-term motivational payoffs. The reason for skepticism is the arguments of Maslow (1953) and of Alderfer (1971) (presented earlier in this article) that when individuals have had ample opportunities to satisfy their social needs, the level of desire for additional social satisfaction will decrease and the level of desire for higher order need satisfaction will increase. When this occurs, of course, then jobs should be high on the four core dimensions (rather than high on an interpersonal dimension) to provide conditions for internal work motivation.

All of the above is not to suggest that the interpersonal aspects of work are unimportant. They do affect satisfaction and it is also possible that they affect other variables which were not considered in this study (e.g., problem-solving effectiveness, turnover). It should also be noted that all of the jobs in the present study were at least moderately high on the interpersonal dimensions; it is very likely that important negative reactions might be obtained from employees on jobs which were very low or totally lacking in interpersonal aspects. Further, the present data address only the degree to which the existing structure of jobs requires or provides opportunities for interpersonal activities. Alderfer (1967) has shown that when jobs are changed, interpersonal relationships (especially between the worker and his supervisor) are markedly affected, and that "relationship problems" clearly have the potential of negating or reversing increases in motivation and satisfaction which are anticipated as a result of job enlargement.

Nature of the Impact of Jobs on Behavior

Until relatively recently, research attention given to jobs has focused mainly on the ways jobs can be designed for most efficient production, or on the analysis of jobs into component parts to facilitate employee selection, placement, and compensation. Since the mid-1950s, however, research attention gradually has shifted toward examination of the effects jobs can have on the people who do them, and how jobs can be designed so that these effects are desirable for both employees and organizations.

Many studies, including this one, have shown that jobs can and often do have a substantial impact on employee behavior and attitudes; this finding is by now rather well documented. It appears that additional understanding of the nature and implications of job effects will require more direct investigation of the process by which these effects take place.

The present study represents some progress in this direction, by providing data which show that certain job characteristics and certain individual differences interact in determining behavior and satisfaction. Apparently, individuals who are desirous of higher order need satisfactions will, when working on a job high on the core dimensions, gradually develop and/or verify the hypothesis that personally valued rewards can be obtained by working hard and effectively on the job. When this occurs, the job can be said to have influenced the behavior of these employees by affecting their personal hypotheses about what kinds of behaviors will lead to favorable outcomes. Thus, one general way that jobs can influence behavior is by creating conditions whereby employees are likely to develop and validate specific behavioral hypotheses about what they can do on the job to obtain work outcomes favorable to themselves (or to avoid unfavorable outcomes).

The general process outlined briefly above represents only one of several possible ways that jobs can influence behavior and attitudes on the job. For example, jobs also can affect behavior by influencing the level of activation of employees at work (Scott, 1966), or by arousing (or depressing) particular employee need states themselves. The general question of how jobs and tasks influence behavior is discussed in some detail by Hackman (1969a; 1970). The point of emphasis here is that job effects can and do come about via a number of different (albeit interrelated) psychological processes, only one of which was given focal attention in the present research. It is likely that sometimes these processes will operate simultaneously and will be mutually reinforcing; other times they may work at cross purposes. Further work toward a more general understanding of the several processes by which jobs affect behavior and how they interact with individual differences would seem to be warranted.

Implications for Organizational Practice

Standard organizational selection and placement procedures attempt to match the skills and abilities of a prospective employee with the skill requirements of the job for which he is being considered. The results of the present research suggest that it may be equally critical for long-term organizational effectiveness to achieve a match between the psychological makeup of the prospective employee and the psychological demands and opportunities of the job. In particular, the present results suggest that individuals who desire higher order need satisfaction will be likely to contribute most effectively to organizational goals (and simultaneously to satisfy their own needs) if they are placed on jobs which are high on the four core dimensions. Other employees, of course, who may be neither desirous of higher order need satisfactions nor capable of dealing with complex jobs requiring considerable autonomy, would be ineffective on such jobs and dissatisfied with them.

It appears from research cited earlier in this article that many organizations err rather consistently by designing jobs which are too low on the core dimensions. The present study supports this conclusion. It suggests that there are many workers who want to obtain more higher order need satisfactions from their work, but few who are overwhelmed by the psychological demands of their jobs. The implication of this argument, of course, is that organizations might be well advised to consider redesigning many of their jobs.

When job enlargement is carried out, the question often arises whether the changes should be toward horizontal enlargement (i.e., increasing the number of different things an employee does) or toward vertical enlargement (i.e., increasing the degree to which an employee is responsible for making most major decisions about his work) or both. Lawler (1969) has reviewed the literature regarding the effects of vertical and horizontal job enlargement and concludes that simultaneous enlargement in both directions may be optimal in most cases. The results of the present study provide some support for this contention. Only if a job is enlarged vertically is an employee likely to feel personally responsible for his work outcomes; and only if a job has some amount of horizontal enlargement is he likely to experience his work as meaningful—although it should be kept in mind that too much horizontal enlargement apparently can cause problems. Simultaneous vertical and horizontal enlargement should increase the likelihood that a redesigned job will be high on all four of the core dimensions.

It should be re-emphasized in conclusion, however, that while jobs appear to be highly potent in determining employee motivation and satisfaction, there is no single best way to design a job. Instead, the results of the present research suggest that the substantial motivational potential of jobs can be realized only when the psychological demands and opportunities of jobs mesh well with the personal needs and goals of employees who work on them. This kind of matching can be developed through selection and placement of employees, through job redesign, or (perhaps optimally) by attempting to fit people to jobs and jobs to people simultaneously and continuously as both the organization and the characteristics of its employees change over time.

REFERENCES

ALDERFER, C. P. An organizational syndrome. *Administrative Science Quarterly*, 1967, **12**, 440–460.

ALDERFER, C. P. An empirical test of a new theory of human needs. *Organizational Behavior and Human Performance*, 1969, **4**, 142–175.

ALDERFER, C. P. *Human needs in organizational settings*. New York: The Free Press of Glencoe, 1971.

ARGYRIS, C. *Integrating the individual and the organization*. New York: Wiley, 1964.

BIGANNE, J. F., & STEWART, P. A. *Job enlargement: A case study*. (Research Series No. 25) State University of Iowa, Bureau of Labor and Management, 1963.

BLAUNER, R. *Alienation and freedom*. Chicago: University of Chicago Press, 1964.

BLOOD, M. R., & HULIN, C. L. Alienation, environmental characteristics, and worker responses. *Journal of Applied Psychology*, 1967, **51**, 284–290.

CONANT, E. H., & KILBRIDGE, M. D. An interdisciplinary analysis of job enlargement: Technology, costs, and behavioral implications. *Industrial and Labor Relations Review*, 1965, **3**, 377–395.

DAVIS, L. E. Job design and productivity: A new approach. *Personnel*, 1957, **33**, 418–429.

DAVIS, L. E., & VALFER, E. S. Intervening responses to changes in supervisor job designs. *Occupational Psychology*, 1965, **39**, 171–189.

DUNNETTE, M. D., CAMPBELL, J. P., & HAKEL, M. D. Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Organizational Behavior and Human Performance*, 1967, **2**, 143–174.

FORD, R. N. *Motivation through the work itself*. New York: American Management Association, 1969.

FRIEDMANN, G. *The anatomy of work*. New York: The Free Press of Glencoe, 1961.

GHISELLI, E. E., & BROWN, C. W. *Personnel and industrial psychology*. New York: McGraw-Hill, 1955.

GUEST, R. H. Men and machines: An assembly-line worker looks at his job. *Personnel*, 1955, **31**, 3–10.

HACKMAN, J. R. Nature of the task as a determiner of job behavior. *Personnel Psychology*, 1969, **22**, 435–444. (a)

HACKMAN, J. R. Toward understanding the role of tasks in behavioral research. *Acta Psychologica*, 1969, **31**, 97–128. (b)

HACKMAN, J. R. Tasks and task performance in research on stress. In J. E. McGrath (Ed.), *Social and psychological factors in stress*. New York: Holt, Rinehart & Winston, 1970.

HALL, D. T., & LAWLER, E. E., III Job design and job pressures as facilitators of professional-organization integration. *Administrative Science Quarterly*, 1970, **15**, 271–281.

HERZBERG, F. *Work and the nature of man*. Cleveland: World, 1966.

HERZBERG, F., MAUSNER, B., & SNYDERMAN, B. *The motivation to work*. New York: Wiley, 1959.

HINTON, B. L. An empirical investigation of the Herzberg methodology and two-factor theory. *Organizational Behavior and Human Performance*, 1968, **3**, 286–309.

HULIN, C. L., & BLOOD, M. R. Job enlargement, individual differences, and worker responses. *Psychological Bulletin*, 1968, **69**, 41–55.

KILBRIDGE, M. D. Reduced costs through job enlargement: A case. *Journal of Business of the University of Chicago*, 1960, **33**, 357–362.

KILBRIDGE, M. D. Turnover, absence, and transfer rates as indicators of employee dissatisfaction with repetitive work. *Industrial and Labor Relations Review*, 1961, **15**, 21–32.

KING, N. A clarification and evaluation of the two-factor theory of job satisfaction. *Psychological Bulletin*, 1970, **74**, 18–31.

LAWLER, E. E., III. Job design and employee motivation. *Personnel Psychology*, 1969, **22**, 426–435.

LAWLER, E. E., III, & HALL, D. T. The relationship of job characteristics to job involvement, satisfaction and intrinsic motivation. *Journal of Applied Psychology*, 1970, **54**, 305–312.

LAWLER, E. E., III, & PORTER, L. W. Antecedent attitudes of effective managerial performance. *Organizational Behavior and Human Performance*, 1967, **2**, 122–142.

LEWIN, K. *The conceptual representation of the measurement of psychological forces*. Durham: Duke University Press, 1938.

LEWIN, K., DEMBO, T., FESTINGER, L., & SEARS, P. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944.

LOCKE, E. A. What is job satisfaction? *Organizational Behavior and Human Performance*, 1969, **4**, 309–336.

LODAHL, T. M., & KEJNER, M. The definition and measurement of job involvement. *Journal of Applied Psychology*, 1965, **49**, 24–33.

LYTLE, C. W. *Job evaluation methods*. New York: Ronald Press, 1946.

MACKINNEY, A. C., WERNIMONT, P. F., & GALITZ, W. O. Has specialization reduced job satisfaction? *Personnel*, 1962, **39**, 8–17.

MASLOW, A. H. A theory of human motivation. *Psychological Review*, 1943, **50**, 370–396.

MASLOW, A. H. *Motivation and personality*. New York: Harper, 1954.

MCGREGOR, D. *The human side of enterprise*. New York: McGraw-Hill, 1960.

PELISSIER, R. F. Successful experience with job design. *Personnel Administration*, 1965, **28**, 12–16.

PORTER, L. W., & LAWLER, E. E., III. *Managerial* *attitudes and performance*. Homewood, Ill.: Irwin, 1968.

SCOTT, W. E. Activation theory and task design. *Organizational Behavior and Human Performance*, 1966, **1**, 3–30.

STIGERS, M. F., & REED, E. G. *The theory and practice of job rating*. New York: McGraw-Hill, 1944.

TOLMAN, E. C. Principles of purposive behavior. In S. Koch (Ed.), *Psychology: A study of a science*. Vol. 2. New York: McGraw-Hill, 1959.

TURNER, A. N., & LAWRENCE, P. R. *Industrial jobs and the worker*. Boston: Harvard University Graduate School of Business Administration, 1965.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

WALKER, C. R. The problem of the repetitive job. *Harvard Business Review*, 1950, **28**, 54–58.

WALKER, C. R., & GUEST, R. H. *The man on the assembly line*. Cambridge: Harvard University Press, 1952.

WEICK, K. E. Laboratory experimentation with organizations. In J. G. March (Ed.), *Handbook of organizations*. Chicago: Rand-McNally, 1965.

(Received March 30, 1970)

JOURNAL OF APPLIED PSYCHOLOGY

AUGUST 1971
VOL. 55, NO. 4

Contents

ARTICLES

Effect of Job Attitudes, Training, and Organization Climate on Performance of the Hard-Core Unemployed: *Frank Friedlander and Stuart Greenberg* 287

Effectiveness of Two Orientation Approaches in Hard-Core Unemployed Turnover and Absenteeism: *Hjalmar Rosen and John Turner* 296

Effect of Individual, Group, and Intergroup Relations on Attitudes toward a Management Development Program: *Clayton P. Alderfer* 302

Motivation in Managerial Levels: Relationship of Need Satisfaction to Job Performance: *John W. Slocum, Jr.* 312

A Study of Entrepreneurial Motivation: Comparison of Fast- and Slow-Progressing Small-Scale Industrial Entrepreneurs in Punjab, India: *P. S. Hundal* 317

Effects of Mode of Organization and Feedback Level on Creative Task Groups: *Thomas H. Stone* 324

Development of the Survey of Work Values: *Stephen Wollack, James G. Goodale, Jan P. Wijting, and Patricia Cain Smith* 331

Environmental Ambiguity and Locus of Control as Interactive Influences on Satisfaction: *Abraham K. Korman* 339

Morale and Productivity Consequences of Group Leadership Style, Stress, and Type of Task: *Leonard L. Rosenbaum and William B. Rosenbaum* 343

Favorable Self-Perception, Perceived Supervisory Style, and Job Satisfaction: *Duane E. Thompson* 349

Influence of Race on the Manifestation of Leadership: *James R. Fenelon and Edwin I. Megargee* 353

Information Seeking in a Natural Stress Situation: *David T. A. Vernon* 359

Identification of Moderator Variables by Discriminant Analysis in a Multipredictable Group Validation Model: *Sheldon Zedeck* 364

General Aptitude Test Battery Scores for Men Differing in Age and Socioeconomic Status: *James L. Fozard and Ronald L. Nuttall* 372

Method for Detecting Random Answers on Personality Questionnaires: *Jerry W. O'Dell* 380

Multiple-Indicant Approach for Studying New Product Adopters: *Jacob Jacoby* 384

Comparative Study of Artists and Nonartists on the Maitland Graves Design Judgment Test: *H. J. Eysenck and M. Castle* 389

Effects of Target Value and Exposure Duration on Recall in a Visual Search Task: *Harvey A. Taub and Charles Abrams* 393

LIST OF MANUSCRIPTS ACCEPTED 379

© 1971 by the American Psychological Association

INFORMATION FOR CONTRIBUTORS

Style. A professional article should possess certain characteristics: (*a*) conciseness and an apparent respect for reader time; (*b*) unambiguous and simple vocabulary with technical and erudite words used only when simpler ones would obviously be inadequate; (*c*) conformity to accepted technical style in tables, terminology, and references; (*d*) conclusions that are clearly related to the evidence presented. The reader should be led, step by step, from a statement of problem or purpose, through analysis of evidence, to conclusions and implications. Authors are encouraged to consult the little book entitled *The Elements of Style* by W. Strunk, Jr. and E. B. White (New York, Macmillan, 1962).

Criteria for evaluation. Manuscripts will be evaluated on the basis of several criteria, including: (*a*) significance in contributing new knowledge to the field, (*b*) technical adequacy, (*c*) appropriateness for the *Journal of Applied Psychology*, and (*d*) clarity of presentation.

Format. Manuscripts must be prepared in the format described in the *Publication Manual of the American Psychological Association* (1967 Revision), obtainable for $1.50 from the American Psychological Association, 1200 17th Street, N.W., Washington, D. C. 20036. Articles not prepared in this manner cannot be reviewed. Special attention should be given to the section on typing the manuscript (p. 48) and to the sections on tables, figures, and references. Note that *all copy must be double spaced*, including references, title, figure captions, etc. The senior author's last name should appear on each page in the upper left-hand corner. In reference lists, give journal titles in full; do not abbreviate.

Copies. All manuscripts must be submitted in triplicate. One copy should be the original typed copy, the other two, clear carbon copies or photo reproductions, only. Authors should check the final typing carefully and retain a copy of the manuscript as a precaution against loss in the mail.

Length. Before preparing a manuscript, the author should check several recent issues to get an idea of the approximate length of regular articles published in the *Journal of Applied Psychology*. (One printed page equals roughly three double-spaced typewritten manuscript pages.) A few longer articles of special significance may be printed from time to time as monographs. Occasionally the Journal will have a section of "Short Notes" featuring brief reports on studies which make some methodological contribution or constitute an important replication.

Abstracts. Each copy of the manuscript must be accompanied by an abstract of 100–120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or they may be obtained from the Editor or from the APA Central Office.

Figures. The original drawing of a figure, or an 8 × 10-in. glossy print of the drawing, is needed for publication and must be submitted with the original typed copy of the manuscript. Duplicate copies of the figure may be photographic or pencil-drawn copies. Figures must be hand-lettered; typewritten lettering is not acceptable.

Early publication. Regular articles are published as nearly as possible in the order in which they were received. Although everything possible is done to keep publication lag to a minimum, at present there is a 10–12-mo. period between the time of receipt and publication of an article. The author of an accepted manuscript may secure early publication, usually in the first issue to go to press after its acceptance, by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication will be sent with the letter of acceptance.

Reprints. No gratis reprints are supplied. Reprints may be ordered from the printer prior to publication.

National Auxiliary Publications Service (NAPS). To save journal space, while at the same time making all necessary materials available to those with a special interest in an article, supplementary materials (such as extensive tables or copies of instruments) may be deposited with the NAPS. Copies of the material to be deposited should be submitted to the Editor with the manuscript and its availability indicated with the following footnote in the article:

> [3] *A five-page table giving data for each of the subgroups* (or other described material) has been deposited with the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

Journal of Applied Psychology
1971, Vol. 55, No. 4, 287–295

EFFECT OF JOB ATTITUDES, TRAINING, AND ORGANIZATION CLIMATE ON PERFORMANCE OF THE HARD-CORE UNEMPLOYED[1]

FRANK FRIEDLANDER[2] AND STUART GREENBERG

School of Management, Case Western Reserve University

The following potential contributors toward the job performance and retention of 478 hard-core unemployed (HCU) were explored: the HCU's biographic and demographic background, his attitudes toward work, the organizational climate in which he is placed, and the effect of a 2-wk. training/orientation program in which he participated. None of the HCU's biographic/demographic background data or his attitudinal characteristics were related to his job performance or retention nor did the 2-wk. orientation program have any effect on his attitudes toward work, job performance, or retention. The sole correlate of the HCU's work effectiveness and behavior was the degree of supportiveness of the organizational climate in which he was placed. In addition, the HCU saw this climate as far less supportive than did his supervisor. An important inference is that programs geared primarily toward adapting the HCU's work attitudes to the predominant social structure in the organization are far less potent than those that also incorporate the adaptation of the organizational climate.

With the sharpened focus directed today toward the training and employment of the hard-core unemployed (HCU), we need a better understanding of the objectives of the methods selected for training and the relative impact of these methods. One of the implicit objectives of such training is to provide the participant from a minority culture[3] with values and attitudes that facilitate his adaptation to the predominant social structure and its culture. The predominant culture is white, employed, more highly educated, of a middle-class socioeconomic level, and one that values highly the ethic of work. The target culture in many training programs is primarily black and unemployed; it has a relatively lower educational and economic level; it values work primarily as instrumental to getting money, rather than for the sake of involvement and having a job (Goodwin, 1969); and its members frequently have a police (or other deviant) record. The emphasis of most training programs is on changing the HCU so that he can adapt to the culture within the organization, rather than attempting to effect change or adaptation in the organizational culture to adapt to the HCU. These cross-cultural issues get acted out by both the predominant and minority cultures within the arena of training programs prior to placement and within the arena of the organization subsequent to placement.

The study that follows dwells primarily on evaluating a program that was intended to recruit and train HCU persons, to provide them with permanent employment in private business organizations, and to support their entry and retention in these organizations after placement. The program in its entirety was directed toward alleviating two interrelated social issues: unemployment, and attitudes inappropriate to the working world of the predominant social structure.

The specific purposes of our research were twofold: first, to evaluate the impact on changes in the HCU's work attitudes (*a*) of a 2-wk. orientation/training program in social skills, and (*b*) of job placement; and second, to examine the predictive power of the following variables as to whether the HCU will become successfully employed: (*a*) his biographic/demographic background, (*b*) his attitudes toward work, (*c*) changes in his attitudes toward work that might result from the training/orientation program, and (*d*) the nature of the job climate in which he was eventually placed for employment. In a larger sense, the authors were interested in exploring the feasibility of increasing the adaptiveness of the HCU to his organizational and social climate; an alternative might have been to attempt to effect change in the climate itself. The variables used in this study to explore these several issues included psychological (biographical data and initial attitudes), social (demographic data), learning (changes in work attitudes as a function of the training program or of having a job), and organizational (the social climate in which the HCU was placed).

---

[1] This study was supported by the United States Department of Labor, Office of Manpower Policy, Evaluation and Research under Contract No. 41-7-002-37. This article does not necessarily represent the Department's official opinion or policy, and the authors are solely responsible for the factual accuracy of all material in the article.

The authors are indebted to Lester D. Jaffe who conducted the interviews from which the climate instrument was derived.

[2] Requests for reprints should be sent to Frank Friedlander, School of Management, Case Western Reserve University, Cleveland, Ohio 44106.

[3] The term "culture" is used in the traditional anthropological sense in that it provides human beings with a design for living, with a ready-made set of solutions for human problems, and so serves a significant adaptive function (Lewis, 1966).

Program and Its Participants

The training and job placement efforts explored in this study were primary components of the AIM-JOBS Program in Cleveland, which was a federally funded Concentrated Employment Program (CEP). The CEP was formulated in 1966 by the President's Commission on Manpower following a Department of Labor report in which some 16–20 cities across the nation were identified as having severe pockets of unemployment and underemployment substantially disproportionate to, and greater than, the remainder of the nation. Of these cities, Cleveland was found to have some of the worst problems of the areas surveyed. AIM-JOBS, the first funded CEP, was established in Cleveland in 1967, with the stated goal, of "placing securely in permanent jobs at least 2,000 disadvantaged residents of inner city poverty areas in the City of Cleveland." An organization of about 140 people was built to achieve these goals through division of work into the areas of recruitment of the HCU, personal assessment, job orientation/training, job development and placement, and supportive services to participants.

The training/orientation program covered a 2-wk. period during which the participant was assigned to a counselor/teacher who worked with the participant individually and in groups. The following subject areas were covered in these sessions: (*a*) human relations—discussions with the broad goal of strengthening respect and understanding for oneself and others through recognition of human diversity; (*b*) job orientation—techniques of getting a job, including how to fill out an application, how to conduct oneself at an interview, acceptable work attitudes, employer expectations, attendance and punctuality, fringe benefits, routes to advancement, etc.; (*c*) money management and budgeting—discussions of the importance of sound money management related to personal goals, budgeting, investment and savings, installment purchasing, credit ratings, etc.; (*d*) grooming and personal hygiene—discussions of proper dress for interviews, personal grooming, etc.; (*e*) physical fitness—including exercise, nutrition, recreation, weight control, etc.; (*f*) transportation and urban development—map reading and use of public transportation systems to reach job sites.

Recruitment into the AIM-JOBS Program was restricted to those who resided in the inner city poverty areas and who were in the age range of 18–35. Applicants were also excluded who had obvious inability to benefit from the Program because of grave physical or mental impairment. Participants in the study were drawn from the 839 HCUs who entered the AIM-JOBS Program during four cycles in the early part of its first year (1967–1968). This group had the following characteristics: 82% male; average age was 26; 91% were Negro; average duration of unemployment prior to job placement was 15 wk.; 20% had prior police records (exclusive of traffic and misdemeanor arrests); average education was completion of tenth grade; average reading skill was approximately at the eighth-grade level; and average arithmetic skill was approximately at the sixth-grade level. In summary, the population for the research was largely male, black, of a relatively low educational level, and obviously unemployed prior to the Program.

Variables

The four major sets of variables included: (*a*) biographic/demographic data, (*b*) mea-

TABLE 1

LISTING OF NINE SCALES OF ATTITUDES TOWARD WORK USED IN STUDY

Attitude scale	r_{tt}	Item	Measure	Source
1. Motivation to work	.66	9	Relative importance of reasons for working	Indik (undated)
2. Motivation to avoid work	.52	7	Degree to which HCU finds work disagreeable	Indik (undated)
3. Motivation to avoid unemployment	.49	4	Discomfort in being unemployed	Indik (undated)
4. Perception of previous job	.71	6	Recollection of previous job	
5. Importance of job characteristics	.64	18	Relative importance of various job characteristics	Friedlander (1965) Wilensky (1964)
6. Prized self-image	.67	6	How HCU would like to be seen by his friends	
7. Powerlessness	.70	6	Degree to which HCU feels he cannot determine his future	Rotter (1966)
8. Cynicism	.73	6	Degree to which HCU is cynical about life and its opportunities	Srole (1956)
9. Vigor	.60	4	Perception of his own pace and vigor	Guion (1966)

Note.—HCU = hard-core unemployed.

sures of the HCU's attitude toward work, (*c*) measures of the organizational climate in which the HCU was placed and employed, and (*d*) criteria of the HCU's job performance.

Biographic/Demographic Variables

A variety of demographic and biographic data was collected from the records of all participants in the Program. Demographic data included length of time at present residence, geographic area of birth, and duration of prior unemployment. Biographic information included number of dependents, number of previous jobs held, education, marital status, prior police record, and level of reading and arithmetic skills.

Attitudes toward Work

A number of measures were administered that were either based on past research or were developed specifically for this study (see Table 1). We hypothesized that scores on the first four scales listed in Table 1 would be predictive of job success and that some of these might be affected by the orientation/training program. For example, the HCU's perception of his previous job experience might well be related to his expectations of this future job experience and, therefore, his job performance and retention. We further hypothesized that the job success of the HCU would be a function of the extent to which job characteristics important to him (Scale 5) were found in his eventual job and the degree to which the role demands of that job were congruent with his prized self-image (Scale 6). Finally, we hypothesized that the HCU's powerlessness, cynacism, and vigor (Scales 7, 8, & 9) would predict attitude changes resulting from the orientation/training program.

The reliability coefficients of these scales are of a minimally acceptable magnitude, although the following two factors may account for this. First, these reliability measures were computed on HCUs in training—that is, the 2-wk. span between test and retest encompassed an orientation program supposedly relevant to these attitudes. In a strict sense, the coefficients are therefore not representative of a pure test-retest reliability, even though no before–after differences were found to be significant. Second, HCUs represent a culture that may well have experienced these questionnaires as alien to its values, resulting in suspicion of and resistance to the questionnaire.

Job Climate

Climate is conceptualized as an interaction of personal factors (personality, needs, values, etc.) and organizational properties (structure, supervisory practices, etc). This relationship emphasizes the role of perception of organizational properties (Forehand & von Gilmer, 1964; Likert, 1961).

Preliminary interviews with HCUs indicated that one of the components of the organizational climate of most concern to the HCU was the degree to which he perceived the climate as supportive. Specifically, three aspects of a supportive climate seemed most salient: (*a*) new worker treatment, (*b*) support from peer workers, and (*c*) support from his supervisor. Both HCUs and their supervisors completed questionnaires concerned with these three aspects of their work climate.

The specific items that comprised each of the three climate measures are listed below. Response options for each item were on a 5-point multiple-choice Likert scale. Items within each of the three measures were combined to yield three climate scores, taking into account the arithmetic sign that follows each item.

1. New worker treatment. (How are new workers at your plant generally treated?) They are usually made to prove themselves (−); They are usually given more breaks than others (+); They are usually treated like all the others (+); They are usually given a hard time (−).

2. Support from peers. (What's it like to work where I work?) Each guy has to pretty much take care of himself (−); Other workers give you a hand and help you if you don't know how to do something (+); Almost everybody gets along well with everybody else (+); Most of the workers are hard to get close to (−); It's not smart to make buddies here because people tend to take advantage of you (−).

3. Support from supervisor. (What's it like to work where I work?) To get ahead, you have to "brown nose" (−); Supervisors would just as soon get rid of you rather than teach you or help you on a job (−).

Criteria of Job Success

Three different criteria of the HCU's job performance were obtained: job retention, work effectiveness, and work behavior. Job retention was considered relevant since one of the major problems claimed in regard to the employment of the HCU was an unusually short duration on the job. Job retention was measured simply by the total number of weeks the HCU worked on the job.

The work-effectiveness criterion was composed of supervisory ratings of four characteristics, each of which was measured by two items: competence (performs his job competently and follows instructions), congeniality (is friendly and agreeable), effort (tries to do his best and works carefully), and reliability (shows up each day, shows up on time). The HCU's supervisor was given the following instructions in this rating: "Compared to other employees doing the same or similar work (or at a similar skill level), how would you rate this employee on each of the following." The multiple-choice format ranged in equal percentile intervals from "top 20%" to "bottom 20%."

The work-behavior criterion was designed to determine the supervisor's general description of the HCU as a person in the work situation. The work-behavior criterion was composed of three component characteristics of the HCU as (*a*) smart (he knows what's going on in life, he does whatever he does well, he knows how to do many things, he is smart); (*b*) friendly (he is a good friend to people, he is a friendly person); (*c*) conscientious (he wants to do his best, he does a careful job, he wants to do a good job). A 5-point multiple-choice Likert-type scale followed each of these items.

METHOD

For questions concerned with evaluating the impact of the 2-wk. orientation program and the impact of working on a job on attitudes and motivation toward work, HCUs completed an identical questionnaire concerned with their attitudes toward work (specified above) at three points in time: (*a*) during the first day that the HCU was involved in the orientation program, (*b*) during the last day of the orientation program (an interval of 2 wk.), and (*c*) after the HCU had been placed on a job and was working (6 mo. after the first administration). The impact of participating in the orientation program on changes in work attitudes was measured by the difference between *b* and *a* above, while the impact of job placement and actually holding a job on changes in work attitudes was measured by the difference between *c* and *b* above.

In addition, participants who held jobs at a point in time 6 mo. after the first administration of the questionnaire also completed the questionnaire on job climate specified earlier. The supervisor of each participant also completed the same job climate questionnaire, and also completed questionnaire forms on the work-effectiveness and work-behavior criteria measures mentioned earlier.

Sample sizes for different parts of the study ranged from 478 to 22. In general, random sampling for each successive phase in which HCUs participated was accomplished, except that those who dropped out were obviously excluded from receiving questionnaires for future phases. Of the 839 HCUs who initially entered the Program during the initial 2 mo. of the study, 478 were randomly selected and completed the first questionnaire on entry into the Program. Biographic/demographic data were available on 300 of these HCU. Of the 428 (90%) who completed the orientation program, 129 were randomly selected and completed the second questionnaire (at the end of the orientation program). Ninety questionnaires were matched on the first two questionnaire administrations. Of the 428 who completed the orientation program, 147 (31%) were working in business organizations 6 mo. later. And of these 147, 110 completed a third questionnaire. Among these 110, 22 were matched with the second questionnaire administration, and 24 were matched with an evaluation by the HCU's supervisor.

A control group, matched by age, education, area of residence, and work history to the AIM-JOBS eligibility requirements, was obtained from other public employment programs. They were given the same questionnaire as AIM-JOBS participants under similar conditions of administration. They were readministered the same questionnaire from 3 to 6 wk. later. A total of 105 matched control group questionnaires was completed.

All analyses used either Pearson correlation coefficients or simple differences between means. Coefficients and differences were tested for significance using a one-tailed t test at the .05 level of significance for correlated or uncorrelated means.

RESULTS

Our findings are divided into two sections: first, those dealing with changes in HCU's attitudes toward work as a result of the orientation/training program and as a result of working on a job and, second, those findings dealing with the prediction of HCU job performance and retention.

Changes in Work Attitudes of the Hard-Core Unemployed

Differences were computed on all attitude-toward-work measures between the first and last day of the orientation program ($N = 90$). None of these differences were significant. Similarly, differences between the HCU's work attitudes on the last day of the orientation program and his attitudes after working on a job (a 6-mo. interval) were also not significant ($N = 22$). Our conclusion is that neither the orientation/training program nor the job experience fostered in the HCU a more adaptive attitude toward work. Indeed, the program had no effect on the HCU's attitudes as measured by the instruments used.

None of the biographic/demographic data was predictive of work effectiveness, work behavior, or job retention ($N = 83$). Furthermore, the biographic/demographic data were not significantly correlated with HCU work attitudes at the outset of the orientation program ($N = 300$) or at a period of 6 mo. later ($N = 83$).

None of the nine measures of the HCU's attitudes toward work at the outset of the orientation program was predictive of work effectiveness, work behavior, or job retention ($N = 83$). The only variable that seemed relevant in any way was the degree of vigor or energy. Those who completed the 2-wk. orientation program ($N = 428$) were significantly higher ($p < .05$) on vigor than those who dropped out before completing the program ($N = 50$).

Our conclusion is that the job performance and retention of the HCU in this study can neither be understood or accounted for by knowing his attitudes and motivation toward work nor by knowing about his previous work history, his geographic mobility, his education and scholastic skills, his family structure, or his prior police record. In a broad sense, the characteristics and perceptions that the HCU brings into the world of work (in terms of the biographic, demographic, and attitudinal dimensions measured in this study) seem to have little to do with whether he performs well on the job and little to do with whether he stays on the job.

Job Climate

This section is concerned primarily with the relationships between the HCU's perception of his work climate and his work effectiveness, work behavior, and job retention. Two findings are relevant to discuss prior to this analysis. One of these is a finding which emerges from a comparison of perceptions by the HCU and his supervisor of the work climate in which the HCU is placed. The second comes from an analysis of the relationships between the supervisor's evaluation of the HCU and job retention.

The dramatic differences between the HCU's and his supervisor's perception of supportiveness of the immediate work climate is illustrated in Table 2. Perceptions by the newly employed HCU of the lack of supportiveness provided to new workers is particularly noticeable. In the case of all three climate variables, the difference in perception is at

TABLE 2

COMPARISON OF PERCEPTIONS OF WORK CLIMATE HELD BY THE HARD-CORE UNEMPLOYED AND THEIR SUPERVISORS

Components of work climate	Perceptions of work climate held by		Difference*
	HCU	Supervisor	
New worker treatment	1.9	4.0	2.1
Support from peer workers	2.5	4.5	2.0
Support from supervisor	2.6	4.9	2.3

Note.—HCU = hard-core unemployed.
* $p < .01$, $N = 24$ (matched pairs of workers and their supervisors).

least two full scale points (on a 5-point scale), and in all cases the critical ratio of the differences exceeds 8. It is apparent that the HCU perceives his work climate as vastly less supportive than does his supervisor.

One of the most surprising results in Table 3 is the high negative relationship ($r = -.60$) between the supervisor's evaluation of the HCU as reliable and the HCU's job retention (number of weeks worked). *Those HCUs who are rated as most reliable* ("he shows up each day," "he shows up on time") by their supervisors *tend to have a relatively short duration on the job.* Or, conversely, those who remain on the job tend to be rated as less reliable by their supervisor.

In general, zero to negative relationships are indicated in Table 3 between supervisory ratings of work effectiveness/work behavior and job retention. It is apparent that these two criteria of job success are measuring, for the most part, two separate dimensions. Perhaps supervisors have little control over retention of HCU persons, this decision is made primarily by the HCU (to quit), or by pressures on the supervisor from higher levels within the organization (not to fire the HCU).

This does not imply that supervisors were, in general, critical of the HCUs working under them. In response to the work effectiveness question in which supervisors were asked to compare the HCU to other employees during the same or similar work, the average percentile ratings for the HCU were as follows ($N = 73$): reliability, 46th percentile; competence, 54th percentile; effort expended, 54th percentile; and sociability, 60th percentile. It should be noted that the HCUs were rated higher on competence than reliability ($p < .05$). When the HCU is present, his work is comparable to that of other employees; the problem is absence and lateness, not his competence.

The main body of findings in Table 3 indicates that HCUs who perceive their climate as supportive also tend to be rated by their supervisor more favorably in terms of work effectiveness and work behavior. Those HCUs who perceive their climate as supportive are consistently rated as more competent and congenial than their fellow workers and as having the general behavioral characteristics of being smart, friendly, and conscientious. HCUs who perceive their climate as supportive

TABLE 3

RELATIONSHIPS BETWEEN PERCEPTIONS HELD BY THE HARD-CORE UNEMPLOYED OF THEIR WORK CLIMATE AND THEIR SUPERVISOR'S EVALUATION OF THEIR WORK PERFORMANCE

Perception of his work climate by the HCU	Supervisor's evaluation of HCU's work performance						
	Comparative rating of work effectiveness				Perception of HCU's behavior		
	Competence	Congeniality	Effort	Reliability	Smart	Friendly	Conscientious
New worker treatment	.49**	.49**	.16	−.03	.43*	.57**	.49**
Support from peer workers	.48**	.55**	.24	−.24	.26	.60**	.40*
Support from supervisor	.61**	.56**	.45*	−.07	.54**	.63**	.58**
No. weeks worked	−.20	−.12	−.12	−.60**	−.23	−.20	−.13

Note.—HCU = hard-core unemployed.
* $p < .05$.
** $p < .01$, $N = 24$.

also show some tendency to be rated as exerting their best effort on the job. The only supervisory rating unrelated to work climate appears in the area of worker reliability, where three negative (but nonsignificant) correlations appear.

Discussion

The effectiveness and behavior of the HCU does not appear to be a function of the personal and social characteristics that he brings into the world of work (such as his motivations and attitudes toward work, his prior work history, geographic mobility, education, scholastic skills, family structure, and prior police record), nor are these characteristics and their resultant behaviors readily susceptible to change through training programs focused on adaptation or enculturation. And, since positive HCU attitudes toward work were uncorrelated with higher job retention or job performance, there is no indication that improved job attitudes (through more powerful social skill training programs) would result in improved job performance or retention.

Rather, the work effectiveness and behavior of the HCU depends predominantly on the social climate in which he is placed and works. Increasing the supportiveness of this climate seems to be a major avenue for increasing the HCU's performance and retention. The implementation of this change lies less in increasing the HCU's competence than in exploring the absence and lateness he shows in response to his climate.

These findings are neither unique to this study nor to the Program studied. Research on JOBS NOW, a Chicago program similar to AIM-JOBS, indicates that participants who were successful and unsuccessful on jobs could not be differentiated on the basis of age, education, IQ, marital status, police record, number of previous jobs, or length of previous employment; the sole factor that differentiated successful from unsuccessful HCU persons at statistically significant levels was the degree of support within the immediate organizational context in which the HCU was placed. The retention rate for "high support" organizations was 82%, while that for "low support" organizations was 28% (JOBS NOW Project, 1967).

The HCU's prime method of coping with a work climate that he perceives as highly unsupportive (and his supervisor perceives as highly supportive) is through his own unreliable behavior—being late and/or absent from work. HCUs who are reliable on the job find the situation intolerable after a short period of time and leave the organization. Others cope with the unfavorable climate by being tardy or absent. Furthermore, our findings indicate that the characteristics that supervisors were most critical of in the HCU was the HCU's lack of reliability, a finding replicated in a number of other companies (Johnson, 1969). Tardiness and absence are characteristics frequently ascribed to the black HCU culture, but these same characteristics can also be viewed as indexes of escape or avoidance behavior from a climate the HCU experiences as intolerable.

The failure of attempts at adaptive training of the HCU should come as no surprise to those familiar with the International Harvester study (Fleishman, 1953). In that study, it was found that after training, foremen's attitudes and values reverted to those dominant in their back-home organizational climate—particularly those of their supervisor. Both studies indicate the impotency of training when the organizational climate does not support values inherent in the training. As Fleishman notes:

> Existing behavior patterns are part of, and are moulded by, the culture of the work situation. In order effectively to produce changes in the foreman's [HCU's] behavior some changes in his "back-home-in the plant" environment would also seem to be necessary. The training course alone cannot do it [p. 215].[4]

The greater the degree to which a program is tied in to company practice and structure, the more effective that program will be. Rosen and Turner (1970), for example, found that an orientation period conducted by members of the hiring organization who were able to intervene between hard-core and first-line supervision in problem areas was significantly more effective than a quasi-therapeutic program focused on individual adaptation issues and conducted by a university.

---

[4] What is in brackets is obviously not Fleishman's, but rather our addition to his statement. It might also be noted that in the Fleishman study, some temporary changes were noted in foreman's attitudes as a result of training, while no such changes occurred in the current study. The assumption is that HCU attitudes are more culturally determined and, therefore, more resistant to training mechanisms.

The supervisor seems to be a key element in the concept of social climate as it effects the HCU. Goodman (1969) comments on the added time demands placed on the supervisor and adds that failure of the supervisor to respond to these demands increases the trainee's dissatisfaction and chances for turnover. Lawrie (1969) notes that the white supervisor often erroneously assumes that the black knows the ropes. When the black's own apprehensions around entry into the white world of work result in tardiness and absence, the supervisor withdraws whatever support he initially provided.

Morgan, Blonsky, and Rosen (1970) found a decreasing acceptance and support by supervisors of an HCU program as it progressed. As causes for this decrease, they mention lack of support of the program by management and a consequent uncertainty by supervisors regarding policy and guidelines in dealing with the HCUs in the program. Hodgson and Brenner (1968) report special supervisory education is desirable where trainees have been identified to supervisors as hard-core before going on the job. Goodman (1969) reiterates this need, specifically in order to reduce the fears and stereotypes held by HCUs and their fellow employees toward each other. In a review of a number of case examples, Dunlop (1969) notes that the attitudes of supervisors, fellow-employees, union representatives, and top management may be as important for successful employment as work attitudes, morale, and incentives of the HCU. Thus, improving the social climate in which the HCU works implies training and ongoing development with all those who compose that climate —managers, supervisors, and peers of the HCU —and not expecting only the HCU to make all the necessary adaptations.

We are suggesting that the very concept of "training" the HCU (by providing him with adaptive skills) places the complete burden for change on the HCU—and eventually on the culture that he represents. Conversely, an adaptive organizational climate implies changes in the relevant others who form and influence that climate—managers, supervisors, and fellow employees—as well as the HCU. This focus suggests a change process based on mutual adaptation, through an integration of attitudes and values of the predominant white middle-class culture and the HCU culture. In an extensive literature review, Amir (1969) found that such an integration process is facilitated (*a*) when there is equal status contact between members of the two groups, (*b*) when the contact is between members of a majority group and higher status members of a minority group, and (*c*) when an "authority" and/or the social climate favor and promote intergroup contact. These findings reiterate the need for programs and interventions that focus on the total organizational system rather than on only one type of person (the HCU) and programs that ascribe equal status to all related parties rather than treat one party (the HCU) as deviant.[5]

Since most HCU programs focus on changing the HCU to adapt to the predominant white middle-class structure, they would seem to proceed on the assumption that the culture of the HCU is defective and that the HCU is accordingly deviant. Wellman (1968), for example, describes a program similar to AIM-JOBS as follows:

> The desire for work was not the problem. The real problem was what the program demanded of the young men. It asked that they change their manner of speech and dress, that they ignore their lack of skills and society's lack of jobs, and that they act as if their arrest records were of no consequence in obtaining a job. It asked, most important, that they pretend *they*, and not society, bore the responsibility for their being unemployed. TIDE [the program] didn't demand much of the men: only that they become white [p. 13].

In a strong criticism of such programs, Valentine (1968) suggests we turn our attention back to our dominant social structure as the source of problems of the poor black, rather than proceed with action programs that would bring the group into "the mainstream of the American way of life" by changing the group rather than the overall social structure. Such traditional action programs keep our attention on the poor black and his "defective culture" rather than on the overall social structure of which we are all a part.

It is within HCU programs such as the one reported here that these cross-cultural issues get acted out by both the dominant culture and the "disadvantaged minority." The results of this study cause us some pessimism with regard to other programs in which the design is intended to change the culturally based attitudes and values of the HCU to fit that of our dominant social structure. Indeed, the social structure as it is represented by the climate in most organizations may be more a cause than a cure for the HCU's inappropriate work behavior.

---

[5] The very term hard-core unemployed connotes deviance of a very durable, chronic type.

REFERENCES

Amir, Y. Contact hypotheses in ethnic relations. *Psychological Bulletin*, 1969, **71**, 319–342.

Dunlop, J. T. Concluding observations. In *Education, training, and employment of the disadvantaged*. New York: National Industrial Conference Board, Inc., 1969. (Studies in Public Affairs, No. 4)

Fleishman, E. A. Leadership climate, human relations training, and supervisory behavior. *Personnel Psychology*, 1953, **6**, 205–222.

Forehand, G., & von Gilmer, B. Environmental variation in studies of organizational behavior. *Psychological Bulletin*, 1964, **62**, 361–382.

Friedlander, F. Comparative work value system. *Personnel Psychology*, 1965, **18**, 1–20.

Goodman, P. S. Hiring, training and retaining the hard-core. *Industrial Relations*, 1969, **9**, 54–66.

Goodwin, L. Work orientations of the underemployed poor: Report on a pilot study. *The Journal of Human Resources*, 1969, **4**, 508–519.

Guion, R. M. The meaning of work and the motivation to work. Paper presented at the meeting of American Psychological Association, New York, September 1966.

Hodgson, J. D., & Brenner, M. H. Successful experience: Training hard-core unemployed. *Harvard Business Review*, 1968, **46**, 148–156.

Indik, B. P. The motivation to work. New Brunswick, N. J.: Institute of Management and Labor Relations, Rutgers, The State University, (undated).

JOBS NOW Project. (Status Rep. No. 3) Chicago, Illinois, Author, 1967.

Johnson, L. A. *Employing the hard-core unemployed*. New York: American Management Association, 1969.

Lawrie, J. W. Making it—The hardest way. *Psychology Today*, 1969, **3**, 29–31.

Lewis, O. The culture of poverty. *Scientific American*, 1966, **215**, 19–25.

Likert, R. *New patterns of management*. New York: McGraw-Hill, 1961.

Morgan, B. S., Blonsky, M. R., & Rosen, H. Employee attitudes toward a hard-core hiring program. *Journal of Applied Psychology*, 1970, **54**, 473–478.

Rosen, H., & Turner, J. Effectiveness of two orientation approaches in hard-core unemployed turnover and absenteeism. *Journal of Applied Psychology*, 1971, **55**, 296–301.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 1966, **80**.

Srole, L. Social integration and certain corollaries: An exploratory study. *American Sociological Review*, 1956, **21**, 709–716.

Valentine, C. A. *Culture and poverty: Critique and counter-proposals*. Chicago: University of Chicago Press, 1968.

Wellman, D. The wrong way to find jobs for Negroes. *Trans-action*, 1968, **5**, 9–18.

Wilensky, H. L. Varieties of work experience. In H. Borow (Ed.), *Man in a world at work*. Boston: Houghton Mifflin, 1964.

(Received April 17, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 296–301

EFFECTIVENESS OF TWO ORIENTATION APPROACHES IN HARD-CORE UNEMPLOYED TURNOVER AND ABSENTEEISM [1]

HJALMAR ROSEN [2] AND JOHN TURNER

Wayne State University

This study is a report of the relative effectiveness of two supportive intervention approaches used with hard-core, unemployed black hires. Effectiveness was judged in terms of both turnover and absenteeism. Rates were compared between the hard-core employees in the two treatment conditions and between hard-core hires and normal hires. The university-sponsored, quasi-therapeutic orientation proved to be less effective than the company-sponsored orientation which, although intended to be a passive, information-giving approach, proved to have a dynamic free-discussion, personal-counseling, and intervention-with-supervision character. Overall, hard-core hires were not differentiated from normal hires in terms of turnover but were characterized by significantly higher absenteeism. It was concluded that extended company-orientation programs could be superior to supportive programs led by outside professionals, and that hard-core hires provided with such orientation would be as stable employees as those hired who met normal hiring criteria.

Industry and business in the United States are increasingly engaging in employment programs for the hard-core unemployed—particularly minority group members from inner city locations of large, industrial urban centers. A good deal of time, money, and effort are being expended to successfully integrate the hard-core into the world of work.

Programs have had, for the most part, one of two emphases: skill training, to compensate for lack of work experience and to develop job competence, and work orientation programs that have focused on developing techniques to aid the hard-core in making an easier and more successful transition from the culture of poverty to the culture of work. The research reported here focuses on the latter type of effort.

As is too often the case when significant policy and practice transitions occur in industry, the pressure to become actively involved minimizes thoughtful and organized preplanning of procedures and strategies. Many companies rushed into such programs on, literally, a trial-and-error basis. As the programs became more popular, however, packaged programs were developed and promoted that implicitly claim to provide the answers to achieving an effective transition for the hard-core (Anonymous, 1968). Unfortunately, there are few, if any, systematic evaluations of the success and/or failure of any of the many orientation efforts. Although many hard-core hiring programs have been given extensive popular coverage in the news media, well-documented statements of program success are almost impossible to find.

The research to be reported is one attempt to provide the needed evaluation of hard-core hiring effectiveness. It should be pointed out, however, that the problems involved in designing and implementing a well-controlled, quasi-experimental design in a field setting are manifest and significantly influence both the results obtained and the interpretations that can be made. Before engaging in a formal presentation of the results of this study, a brief discussion of problems encountered and their impact is necessary.

One of the most critical problems encountered was related to the matter of implementing the controls demanded by the design in an ongoing, industrial organization. In this study, the intention was to test the relative utility of a quasi-therapeutic approach to job orientation of the hard-core. This university-conducted orientation was to be contrasted to the impact of an equally extensive, passive, information-centered orientation carried out by company training personnel. Only after the program was well underway was there a realization that the informational orientation had become far more than intended: Discussion was encouraged using the information presentation as a springboard. Individual counseling of problem employees, designated as such by line supervision, occurred frequently. And, in many cases, training personnel interceded on the behalf of the employee in the relationship between first-line supervision and the hard-core employee. What had been designated as a passive placebo appeared to have become a dynamic, personalized treatment condition. As a consequence, this study became less a test of the impact of quasi-therapeutic style orientation per se and more a study of how effective a company-run orientation could be, given commitment, latitude, and innovation on the part of the training staff.

---

[1] This research was supported in part by Labor Department Contract 81246824.

[2] Requests for reprints should be sent to Hjalmar Rosen, Department of Psychology, Wayne State University, Detroit, Michigan 48202.

Another critical problem encountered, and one characteristic of many research efforts where intervention is used (Webb, Campbell, Schwartz, & Sechrest, 1966), relates to the fact that any modification of an organizational environment has the potential for being reactive. Bringing hard-core into a plant as special employees, requiring extensive off-the-job support time for them and their supervisors, well may have communicated to first-line supervision the company's commitment to "make the program work at any cost." Absenteeism data (presented in Tables 2, 3, and 4) suggests that a leniency occurred regarding absences for the hard-core in comparison to the normal hire. Consequently, differences between hard-core and normal hires probably underrepresents the true magnitude of the difference—particularly with regard to turnover via termination.

One final limitation should be noted. This study focused on job-stability characteristics of hard-core. Any attempt to generalize from these findings to results of other studies and programs is hazardous. Not only would generalization be limited by lack of comparability in intervention strategies imposed, but it also would be limited by lack of comparability among the focal groups. Unfortunately, in many programs, the hard-core designation refers to urban, minority group members—usually black—unemployed at the time of hire, that is, persons that company personnel officers privately confide are not really different from the general run of new hires in a tight labor market. Hard-core in this study refers to persons having long histories of unemployment for a variety of reasons (some of which relate to being black), and therefore represents a difficult group to retain in employment. To compensate for this lack of comparability, job stability was compared not only between the two intervention groups of hard-core but between those groups and normal hires employed during approximately the same period and assigned to similar, unskilled positions.

METHOD

Subjects. Forty-nine Negro *S*s (males), who had been residents of the urban community in which the sponsoring company was located for at least 1 yr., made up the hard-core sample. Hard-core unemployment was defined essentially in terms of work history. To be permitted to enter the program, the applicants could not have held any single job for more than 3 mo. and had to have total employment of less than 6 mo. during the 2-yr. period prior to hiring. Educational level, selection test data, and criminal record (except for crimes of extreme violence) were ignored. The *S*s did, however, have to be evaluated as physically able to work by the estimation of the company medical staff. The *S*s were obtained from welfare roles, supportive agencies, and from day-labor lines. Normal hires were defined as walk-in applicants who met the typical hiring standards of the company. This second group happened to be one-third black and, like the hard-core, were assigned to manual, unskilled level jobs.

Setting. The employing company was a public utility. The company was located in a large, Midwest urban center that, during the year prior to the study, had experienced severe race riots. A large proportion of company employees were technicians rather than unskilled workers. Most new nonprofessional or clerical employees, however, were assigned to an unskilled labor pool. The company, as a matter of policy, viewed the labor pool as a resevoir for future technical personnel rather than as a stable, unskilled work force. It actively encouraged job bidding and training for higher skill positions. At the time of the study approximately 10% of the work force was black.

Procedure. The hard-core sample was divided into two groups. Group designation was determined by matching in terms of demographic variables obtained during the hiring interview.[3] Orientation for both groups consisted of 21 hr. of exposure in 14 sessions over a 12-wk. period.[4]

---

[3] Critical variables included age, education, number of dependents, and criminal record. As far as possible, job assignments were balanced so that any given supervisor would have an *S* from each type of orientation group assigned to him.

[4] For the first 4 wk., two 1½-hr. sessions were held. For the next 4 wk., there was one session per week.

TABLE 1

TURNOVER RATE AND AVERAGE LENGTH OF EMPLOYMENT BEFORE TERMINATION FOR NORMAL HIRES AND UNIVERSITY- AND COMPANY-TRAINED HARD-CORE

Measure	Normal	Hard-core	University hard-core	Company hard-core
Original N	49	49	24	25
No. terminations	12	16	11	5
% Terminations	24.49	32.7	45.83	20.0
Average no. days before termination (5-day week)	59.15		65.45	56.25

University-developed training was conducted with three groups of eight *S*s each. The training style used a quasi-therapeutic approach. Although the direction was in terms of job-related problems, complete freedom was permitted in group meetings. The groups were led by a white, clinical psychologist who had had considerable experience in group psychotherapy with mixed racial groups.

The company-developed training was directed by a white member of the personnel division of the company. The program content was developed by the training division of the company from existing orientation materials and primarily stressed job-related information. The company-trained hard-core met as a single group of 25 *S*s. Although it was intended that company orientation be a passive treatment—that is, essentially a lecture, tour, and film series—it was discovered that considerable free discussion was permitted during the group sessions. Moreover, toward the end of the training sessions, the group leader used this time to talk to *S*s who were having job-related problems. Neither modification was planned in the study design.

As noted earlier, three types of criterion data were collected. After the tenth week of the program, immediate supervision was requested to evaluate hard-core employees in terms of recommendations for a wage increment.[5] At the end of 6 mo., turnover data and absenteeism data (both authorized and unauthorized) were collected from company personnel records and tabulated for both the hard-core and the normal-hire groups.

RESULTS

At the tenth week after hiring, 44 hard-core remained employed. In terms of supervisor evaluation for a pay increment, 57% of the hard-core were considered worthy of the merit raise. The company-trained group had 65% within this category and the university-trained group had 48%. Although the advantage was not statistically significant, it was the first indication in the study of the possible superiority of the company-sponsored training over that offered by the university.

At the end of 6 mo., 34 *S*s, or 69% of the original hard-core hires, remained on the company employment rolls. At this point, turnover data for the two hard-core training groups and the normal-hire sample were compared. From Table 1, it can be noted that the hard-core as a composite had a turnover rate, which, although higher, was not significantly different from that of the normal-hire group. The university-trained group had a significantly higher ($p = .01$) turnover rate than did either the company-trained group or the regular hires. Termination time, approximately 12 wk., did not differentiate among the three groups.

In comparing the three groups on 6-mo. authorized and unauthorized absence rate (computed in terms of percentage of lost time to total time worked), no significant differences were found in terms of authorized absence (see Table 2). The university-trained group had the highest rate. In terms of unauthorized absence rate, however, significant differences occurred. Both hard-core groups had a significantly higher rate ($p = .01$) than did normal hires. Moreover, the university-trained group had a significantly higher ($p = .05$) rate than did the company-trained group.

Breaking down the absenteeism data in terms of stays and leaves for the three groups, Table 3

TABLE 2

COMPARISON OF AUTHORIZED AND UNAUTHORIZED ABSENTEEISM RATE FOR TOTAL SAMPLES OF NORMAL HIRES, UNIVERSITY-TRAINED HARD-CORE, AND COMPANY-TRAINED HARD-CORE

Measure	Normal hires[a]	University trained[b]	Company trained[c]
% Authorized	3.52	6.67	4.43
% Unauthorized	2.20	11.83	6.48

Note.—Comparison in terms of percentage of time lost.
[a] $N = 49$.
[b] $N = 24$.
[c] $N = 25$.

During the final 4 wk., a session was held every other week. All sessions were conducted on company time and the employees were paid their normal hourly rate.

[5] The company had a policy that permitted a small merit increment for employees who were considered to have good potential on the basis of recommendations from their immediate supervisors.

TABLE 3

COMPARISON OF AUTHORIZED AND UNAUTHORIZED ABSENTEEISM RATE FOR THOSE REMAINING EMPLOYED AFTER SIX MONTHS

Measure	Normal hires[a]	University trained[b]	Company trained[c]
% Authorized	.71	7.93	4.85
% Unauthorized	1.95	6.79	4.44

Note.—Comparison is in terms of percentage of time lost.
[a] $N = 37$.
[b] $N = 13$.
[c] $N = 20$.

TABLE 4

COMPARISON OF AUTHORIZED AND UNAUTHORIZED ABSENTEEISM RATE FOR THOSE TERMINATED DURING THE FIRST SIX MONTHS

Measure	Normal hires[a]	University trained[b]	Company trained[c]
% Authorized	7.88	6.53	3.80
% Unauthorized	6.33	16.44	12.99

Note.—Comparison is in terms of percentage of time lost.
[a] $N = 12$.
[b] $N = 11$.
[c] $N = 5$.

presents the data for the stays, that is, for those remaining employed after 6 mo. The rates of authorized and unauthorized absences were higher for the university-trained group than for the company-trained group, although not significantly so. Both hard-core trained groups had significantly higher ($p = .01$) unauthorized absence than did the normal hires, and both were higher, but not significantly so in authorized absence.

The trends were much the same for the terminated employees, although the absenteeism rate was considerably higher than for the nonterminated employees (see Table 4). University-trained and company-trained hard-core were not significantly differentiated from one another, either in terms of authorized or unauthorized absence. In terms of these variables, however, the hard-core groups did have significantly higher ($p = .01$) unauthorized rates than did the normal hires. No significant differences were found among the three groups regarding authorized absence.

To estimate whether or not absenteeism played a role in termination,[6] the rate of absenteeism for stays versus leaves in the three groups were analyzed (see Table 5). For all groups, unauthorized absenteeism was significantly higher for the leaves than the stays ($p = .01$). For the terminated employees in the hard-core training groups, the rate of unauthorized absence was significantly higher ($p = .01$) than authorized absence.

DISCUSSION

Keeping in mind the variance from the intended design, particularly with regard to the company-sponsored orientation, two important aspects of the results stand out. The first relates to the comparability between the hard-core (taken as a totality) and regular hires in terms of turnover, in spite of the significantly higher absenteeism rates for the hard-core. The second is the clear advantage in terms of all of the criteria used of the company-trained versus the university-trained hard-core hires.

In terms of the hard-core versus the normal-hire findings, one could summarize the impli-

TABLE 5

COMPARISON OF UNAUTHORIZED AND AUTHORIZED ABSENTEEISM RATE FOR STAYS VERSUS LEAVES IN NORMAL HIRES, UNIVERSITY-TRAINED, AND COMPANY-TRAINED GROUPS

Measure	Normal hires		Total hard-core		University trained		Company trained	
	Stays	Leaves	Stays	Leaves	Stays	Leaves	Stays	Leaves
% Unauthorized	.71	6.33*	6.44	15.36*	7.93	16.44*	4.85	12.99*
% Authorized	.95	7.88*	5.62	5.67	6.79	6.53	4.44	3.80

Note.—Comparison in terms of time lost.
* $p = .01$.

---

[6] From other data in the study, that is, exit interviews and foreman training session tapes, it appeared that absenteeism was a significant variable in termination. Company records indicating formal reasons for termination supported this conclusion.

cations of the results in the following manner: If hard-core unemployed are given extensive orientation training, particularly if such training is job oriented in content and directed by company personnel, and if company standards regarding absenteeism are loosened during the early employment period, the hard-core will turn out to be as stable a work force as men hired via normal hiring channels.

The key to developing job stability among hard-core employees, then, seems to be a matter of modifying poor work habits and developing an acceptance (and understanding) of the rigorous and systematic attendance demands made by a large business organization on its work force. Habit patterns developed in extended periods of unemployment, particularly related to time, apparently have to be extinguished before the hard-core hire can effectively operate within the constraints a steady job imposes on his personal freedom. This change, for many, will not occur spontaneously, and must be carefully nurtured over a considerable period of time. The comparisons of the two hard-core orientation approaches shed additional light on how this adaptation might best be brought about.

The only apparent advantage the university orientation had, as compared to company-directed orientation as it actually was carried out, related to the professional stature and experience of the orientation leader in using quasi-therapeutic, participative group techniques. The quasi-therapeutic approach, in and of itself, could not a priori be evaluated as a superior acculturation strategy over the free discussion, personal counseling, and intervention with supervision used in the company-orientation procedures.

On the other hand, the company strategy as it was implemented had some obvious advantages over the university approach. First, the subject matter covered in the sessions was job oriented and crucial to acceptable work behavior. Benefits derived from working for the company were spelled out and relevant rules and penalties for their infraction were empha- ized time and again. The rationale for the ules and penalties also was discussed at length. n contrast, the university trainees established heir own content, much of which was not di- ctly related to their jobs—or, for that matter, to work in general—but rather to a protest against white establishments and life in general. Work problems tended to be brushed aside with the most cursory discussion in spite of the efforts of the trainer. Only in terminal sessions did the groups begin to focus on job problems.

A second accrual of the company trainees was the direct access the training supervisor had to the organization and the fact that he represented the company. He knew the rules and could assess them realistically, particularly in terms of the supervisors that would be implementing them. Given a problem employee, he not only could call him aside after the meeting and give him a clear picture of the consequences related to rule breaking, but could and did intercede with the employee's supervisor on the employee's behalf. In a sense, company training included a liaison function that was absent from the university-orientation training and may have played a critical role.

The poor showing of the university-sponsored, quasi-therapeutic group orientation raises several critical questions. First, is it advisable to have orientation initiated by a source outside the company organization? Second, is there utility in using a clinically-derived adjustment technique for hard-core employees in terms of the critical job-behavior criteria used, that is, supervisor ratings, turnover, and absenteeism?

Assuming for the moment the need for special intervention in the form of orientation, the outsider is at a distinct disadvantage regardless of his professional stature and experience. He is not apt to be well-versed in company practices and policies. Without elaborate exposure and probing, the most he will obtain is a set of formal statements undifferentiated in terms of importance and enforcement. Second, he is not in a position to intercede for the employee either to clarify rule-enforcement standards of supervision with regard to a given infraction or to request a temporary, less rigorous rule enforcement.

On the basis of such logic, there is also a very real question as to the merit of using a quasi-therapeutic group approach. The goals of such a strategy, on one level, make a good deal of conceptual sense. Sharing of common problems should not only be supportive but should lead to a more realistic group-derived solution. In turn, assuming that the problems are job related—even if indirectly so—the trainee should benefit by being able to cope with his adjustment problems more successfully. Theoretically, he should be less likely to quit because of the group support and should learn to modify his behavior and thereby reduce punitive action from the organization.

A critical assumption, however, is that job-related problems will be salient ones for him and will be the problems he and his group will discuss. This assumption did not hold true for the group studied. Job-related problems were lost in a mosaic of problems induced by being black and living in the ghetto.

Although, ideally, personal adjustment may be desirable for job success, its development may be too amorphous and gradual to help in coping with ongoing, job-related problems. To remain employed, the employee, hard-core or not, must conform to avoid dismissal and must accept to minimize self-termination. These adjustments must be achieved in a relatively short period of time regardless of his overall personal adjustment. The quasi-psychotherapeutic group treatment, as implemented, did not seem to achieve these ends.

Summary and Conclusion

In terms of the findings of this study and the conditions under which they were generated, there is little doubt that hard-core orientation can be advantageously accomplished, in terms of the criteria used, by: (*a*) introducing a content structure in orientation that focuses on work-related materials, (*b*) having the program administered and conducted by members of the hiring organization who are not only knowledgeable about company practices and policies but who can intervene between hard-core and first-line supervision in problem areas, and (*c*) using a part of the orientation period to cope with and counsel problem employees. Extra-organizational intervention, particularly using a quasi-therapeutic group strategy, does not seem to be an optimal or viable alternative.

REFERENCES

Anonymous. How to succeed in hard-core hiring. *Business Week*, August 24, 1968, 65–66.

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. *Unobtrusive measures.* Chicago: Rand McNally, 1966.

(Received April 8, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 302–311

EFFECT OF INDIVIDUAL, GROUP, AND INTERGROUP RELATIONS ON ATTITUDES TOWARD A MANAGEMENT DEVELOPMENT PROGRAM[1]

CLAYTON P. ALDERFER[2]

Department of Administrative Science, Yale University

The present study describes the impact of a management development program in a medium-sized bank that was designed to increase the number of competent and aggressive young managers available to the organization. In this situation after a period of 10 yr., the training program had developed into a subsystem of the organization. One's attitudes toward the program were a function of his seniority, organizational rank, pay satisfaction, and perception of his own career prospects. The kinds of attitudes one held about the program also depended heavily on his status as an employee, trainee, or officer, and on his sex.

Almost any formal organization must continually face the problem of providing a competent corps of managers to fill positions at the middle and upper levels of the system. Most organizations deal with this problem by a combination of recruitment and management development. From outside the system, the organization attempts to attract people whom they imagine will be able to fill present and future management positions. From among current staff, both management and nonmanagement, they attempt to train or in other ways prepare people for positions of greater responsibility in the future.

At the intersection of selection and training, the theoretical orientations of administrative theory and industrial psychology coincide. Both approaches tend to view selection and training as partially interchangeable functions that serve to provide organizations with individuals who are properly prepared to meet the demands of specific roles (Dunnette, 1966; Etzioni, 1964; Simon, 1957). Both individuals and organizational roles need to be understood and adjusted to attain a satisfactory fit between the individual and the organization. However, if analysis stops at this point, group and organizational level factors that bear on management development tend to be ignored.

From the standpoint of scientific parsimony, there may be no need to consider more complex human groupings if the individual level of analysis provides adequate conceptual and methodological tools to account for observable phenomena. But if the individual level of analysis is not adequate, then more elaborate conceptual schemes are necessary.

Open Systems Theory

The concepts of open systems theory offer an important approach to thinking about the interrelationship among human units at different levels of analysis (Argyris, 1960; Katz & Kahn, 1966; Rice, 1969, Schein, 1965). All units have boundaries that serve to define their existence and differentiate them from other parts of a total system. The concept of an open system has been developed because no living human system is completely isolated from other systems outside its boundary. All human systems and subsystems exchange matter, energy, and information with their environments (Emery & Trist, 1965).

The existence of a management recruitment and development program may be viewed as an organization coping with its needs for survival and growth by setting up a subsystem within the organization to attract managers and prepare them for future responsibilities. Upon its creation, such a program defines a new group within the organization and a new way for the system to relate to its environment. This trainee group also has relationships with other groups in the organization. Initially, a trainee becomes a member of the general male or female employee groups, but most often he is a temporary member, shortly to become an officer of the bank. Until, and sometimes after, he becomes a member of management, the trainee may be viewed as a partial or temporary member of the officer corps (Glickman, Hahn, Fleishman, & Baxter, 1968).

---

[1] The author would like to thank Robert Kaplan, Chris Argyris, Richard Hackman, Charleen Alderfer, and all the members of the organization for their help and cooperation with this study.

[2] Requests for reprints should be sent to Clayton P. Alderfer, Department of Administrative Science, Yale University, 2 Hillhouse Avenue, New Haven, Connecticut 06520.

Exclusively individual level analyses of management development would not identify these various group boundaries that a trainee crosses during his preparation for management responsibilities. The function of boundaries in identifying, organizing, and protecting systems and subsystems means that their crossing cannot be taken lightly. It nearly always signals potential threat (Rice, 1969). One would, therefore, hypothesize that the reactions of organization members to a management development program would be a function not only of individual level variables but also of group level variables.

The objectives of this study were: (*a*) to test the foregoing hypothesis, (*b*) to explore ways to enrich and elaborate the general hypothesis, and (*c*) to follow up a study by Argyris (1954) in the same bank.

Fifteen years earlier Argyris had identified a tendency for the organization to select, develop, and retain a certain type of employee. He coined the term "right type" to characterize the preferences and behaviors of this employee, who tended to be quiet, slightly nervous, very willing to take orders, and relatively unwilling to give orders. This person tended to have preferences for security, stability, predictability, isolation, and the avoidance of aggression in himself and in others. Since Argyris' study, a number of changes occurred that might have altered the predominance of "right type" employees. A number of mergers were effected, new services were added, and they tended to draw people from other industries than banking, such as finance companies, sales and advertising agencies, and the computer-processing field. In addition to these institutional changes, approximately 10 yr. prior to the study the organization instituted a management development program. The purpose was to attract and help develop a corps of highly intelligent and aggressive men prepared to staff the middle and upper levels of management positions in the bank.

Operation of the Program

The actual administration of the program was learned about gradually during the study. The procedures were common practice, but they were not common knowledge except to the few people who had direct responsibility for administering the program. There was, for example, no written document outlining the procedures. Initial screening of applicants for the program, often through interviews on college campuses, was done primarily by two of the first members of the training program who were in their early 30s. After initial screening, some candidates were invited to visit the bank for further interviews. At the main office, they faced a series of interviews that always included the President, the Head of Personnel, the Executive Vice-President, and several Senior Vice-Presidents. After the day of interviews, the interviewers discussed their impressions of the candidate, sometimes together and sometimes by being polled by one of the program administrators. Usually the decision of whether an employment offer was to be made was communicated by the man who had done the initial screening of the applicant. If a person received and accepted an offer, he began a 33-wk. program containing several components. First, he was to spend some time in most of the major departments of the bank, learning their respective operations. Each department designed its own component of the program. During the 33-wk. period, each trainee was assigned a "big brother," who was a graduate of the training program. The big brother's role was to aid the smooth functioning of the junior man. He helped schedule where and when the trainee would be; he shared whatever information he had about the respective departments; and he saw key people in the various departments to obtain an evaluation of his trainee's performance in the departments. It occasionally happened that trainees were terminated at the end of the training period if their evaluations wer uniformly poor. Otherwise, a trainee woul then be assigned to a full-time position. I usually took at least 18 mo. until a trainee wa made a member of management after he ha completed the orientation phase.

METHOD

Procedure

Five different approaches to data collection were used to provide information about the management development program: (*a*) individual interviews designed to cover a wide range of work attitudes, (*b*) a fixed alternative questionnaire covering a number of satisfaction issues, (*c*) group interviews using the individual interview questions that were appropriate to ask in a group setting, (*d*) a fixed alternative questionnaire asking specifically about the management training program, and (*e*) group feedback sessions in which the results of the preceding data were presented and discussed with interested members of the organization.

The individual interviews, general satisfaction questionnaires, and group interviews were used concurrently in the first phase of the study. Individuals were selected at random to participate. At the conclusion of the individual interview each person was given the satisfaction questionnaire and asked to return it by mail. Each person was assured that his answers in both settings would remain confidential. Group interviews were conducted with groups of 6–12 people. The groups were composed so that management and employees would not be in the same session. Other than this constraint, assignment to interviews was on a random basis. The participants were also asked to agree among themselves to refrain from discussing what happened in the interview after it was over. Group interviews provide two opportunities that were not found in individual interviews. First, they provide a setting in which the participants can stimulate each other in ways that are not possible for an investigator working alone. Second, they provide an opportunity for the investigator to observe actual behavior of organization members in relation to one another. In this way, the group interview provides a microcosm of the organization in a quasi-controlled setting.

At both individual and group interviews, questions like, "How does one get promoted in this organization?" or "Is there anything else that a person should understand if he is to capture the essence of this organization?" led respondents to discuss the management development program. In addition people chose to write in comments on the questionnaire, indicating that they would like to have an opportunity to answer questions about the program.

In response to these reactions a number of additional methodological tactics were employed. A question specifically on the program was added to the group and individual interviews, and a special questionnaire focusing on the program was designed and mailed to all employees in the organization. This questionnaire primarily consisted of a series of 24 statements that people had made about the program arranged in a Likert-scale format. Finally, group interviews with trainees only were arranged so that these men might talk together about their experiences with the program.

The final source of information about the program came from the sessions that were held with the officers of the bank in order to report to them what had been learned from the organizational study. After all of the data had been collected and analyzed, two reports were prepared for the organization. One focused specifically on the management development program, and the other covered the organization as a whole. After the top management group had reviewed these findings, together with the investigator, they decided to make the reports available to any officers who wished to read them, provided that these people would then attend some group discussion meetings after they had read the reports. These groups consisted of 6–8 people who shared common organization positions. No one attended a session in which his immediate supervisor was present. These meetings provided an opportunity to enlist the members of the organization in explaining the various findings uncovered by the study.

Response Rate

At the time of the study, the bank employed approximately 700 people of whom approximately 120 were officers (the title given to a person when he was made a member of the management team). About $\frac{2}{3}$ of the employees were women, while all but five of the officers were men. The term trainee was used for any person who was or had been a member of the management development program. Twenty-five trainees were employed by the bank, 11 of whom were also officers.

The management development questionnaire asked people to return the instrument even if they did not feel enough familiar with the program to answer questions about it. Twenty-two trainees, 80 officers (in addition to those trainees who were officers), 62 male employees, and 99 female employees returned useable questionnaires. One trainee, 7 officers, 27 male employees, and 115 female employees returned the questionnaire without answering the attitude items.

RESULTS

Factor Analysis of Attitudes

The 24 attitude statements about the program were factor analyzed by the principal components method. All factors with Eigenvalues greater than one were rotated according to the Varimax procedure. In addition to the statements explicitly about the program, four other statements were included in the questionnaire to measure the person's perception of his career promise and his pay satisfaction. These items were factor analyzed along with the items about the program.

The results of the factor analysis and rotation produced six interpretable factors about the program and two additional factors about pay satisfaction and career prospects as shown in Table 1. All of the items on the questionnaire resulted in loadings of .40 or higher on a factor. The first factor might be characterized as

TABLE 1

PRINCIPAL COMPONENTS FACTOR ANALYSIS AND VARIMAX ROTATION OF STATEMENTS FROM MANAGEMENT DEVELOPMENT QUESTIONNAIRE

Statement	Factor							
	1	2	3	4	5	6	A	B
1. The Management Development Program is making an important contribution to the management of the bank.	−.78	−.09	.10	−.08	−.16	−.01	.04	−.09
2. The bank should give higher pay and responsibility to trainees if necessary to keep them from being drawn away by more attractive positions with other organizations.	*−.64*	−.14	.00	.07	−.01	.33	−.07	.13
3. The Management Development Program has made the bank a more attractive place for me to work.	*−.56*	−.19	−.02	−.16	−.25	.33	−.14	−.10
4. On the whole, the trainees are a bright, competent group.	*−.63*	.07	.17	−.14	−.27	.00	.08	−.30
5. The Management Development Program does not provide effective training.	*.55*	.26	−.13	.03	.07	.01	.15	.18
6. Trainees are being given positions of responsibility before they have enough experience to be able to handle them effectively.	*.53*	.44	−.22	−.02	.08	−.13	.32	.11
7. The Management Development Program has outlived its usefulness.	*.71*	.28	−.02	.19	.05	.09	−.01	.06
8. Trainees tend to have less loyality to the bank than others.	.29	*.73*	−.13	.05	.12	−.03	−.18	.14
9. Trainees are given undeserved preferential treatment.	.50	*.57*	−.14	.17	.15	−.16	.14	.13
10. Once a person is selected for the Management Development Program, he is almost guaranteed to move up the ladder rapidly in the bank.	.32	*.44*	.42	.38	.11	−.17	.28	.16
11. In the future, it will be increasingly difficult for nontrainees to reach positions of senior responsibility in the bank.	.05	*.49*	−.21	.45	.02	−.10	.17	.10
12. Trainees are more willing to use the bank for their own personal gain than other organization members.	.15	*.68*	−.18	.19	.13	−.08	−.10	.16
13. Some trainees have a great deal of influence with senior management—more than their positions justify.	.13	*.62*	−.21	−.02	.16	.13	.24	.18
14. There is a tendency for trainees to have closer relationships with each other than with others in the bank.	.12	.32	*−.79*	.05	.10	.04	−.06	−.05
15. Trainees tend to associate mainly with other trainees.	.12	.31	*−.80*	.04	.04	−.03	−.02	.05
16. The standards for evaluation by the bank are higher for trainees than for others.	−.16	.24	*.44*	−.05	−.13	.23	.27	.12
17. The Management Development Program has made it harder for me personally to progress in my career with the bank.	.32	.13	−.09	*.47*	.07	.10	.05	.45
18. There should be a clearly specified way for current members of the organization to be considered for the Management Development Program.	−.08	.15	−.41	*.54*	−.05	−.06	.11	.01
19. Currently there is no opportunity for present employees to be considered for the Management Development Program.	.31	.08	.12	*.72*	.27	−.06	.06	.04
20. Most trainees are graduates of Ivy League colleges.	.12	.13	−.16	.02	*.87*	.04	.00	.06
21. If a person has graduated from an Ivy League college, he doesn't stand much of a chance of being considered for the Management Development Program.	.27	.30	.01	.24	*.68*	−.05	.15	.16
22. Trainees are not given positions of responsibility soon enough in their careers.	−.14	−.04	.24	.13	−.03	*.63*	−.15	−.19
23. The Management Development Program has directly or indirectly made it easier for me personally to progress in my career with the bank.	−.42	−.06	.00	−.02	−.15	*.58*	−.26	.13
24. Sometimes trainees are not promoted as fast as they might be in order not to upset other trainees.	.10	−.02	−.08	−.22	.12	*.81*	.14	−.01
1. I feel that I have a very promising future with the bank.	−.11	.09	−.03	−.24	−.10	−.02	*−.81*	−.19
2. I feel that my own career in the bank has just about reached its peak.	.04	.10	.10	.01	.00	−.13	*.85*	−.11
1. I feel that my compensation is quite satisfactory.	−.08	−.17	.05	.03	−.06	.14	−.02	*−.85*
2. Right now I feel underpaid for my work.	.17	.14	.04	.14	.08	.02	.04	*.85*

Note.—Each statement was answered on a 6-point scale in which 1 = strongly agree and 6 = strongly disagree. The italicized numerals indicate which items were summed to form attitude scales. The last four items contain the pay and future prospects scales which dealt indirectly with attitudes toward the program.

dealing with an overall evaluation of the program. Items from this factor dealt with the overall contribution of the program to the bank's management and with the training effectiveness of the program. The second factor had to do with the degree to which the trainees received undeserved preferential treatment and misused their privileges. There was a close association between seeing the trainees as receiving undeserved preferential treatment and seeing them as abusing their status. A person who saw the trainees as receiving special privileges also saw them as misusing their influence and vice versa. The third factor concerned the degree to which the trainees were perceived as a closed cliquish group. Also associated with this factor was an item dealing with the degree to which the trainees were seer as having to measure up to higher standards o evaluation than others in the bank. The fourt

TABLE 2

CORRELATIONS BETWEEN INDIVIDUAL LEVEL VARIABLES AND REACTIONS TO THE PROGRAM

Reactions to program	Individual satisfactions and roles			
	Promising career	Pay satisfaction	Seniority	Rank
Evaluation	.20**	.32**	−.15*	.12
Preferential treatment	−.21**	−.37**	.21**	−.04
Closed group	.13	−.08	.17*	.26**
Career barrier	−.20**	−.34**	−.05	−.34**
Ivy League graduates	−.15*	−.25**	.11	−.11
Accelerate trainees	.18*	.10	−.14*	.05

Note.—$N = 263$.
* $p < .05$ (two-tailed test).
** $p < .01$ (two-tailed test).

factor concerned the degree to which the program was perceived to be a barrier to one's own career development. The fifth factor concerned the degree to which trainees were seen as exclusively graduates of Ivy League colleges. The sixth factor dealt with the degree to which the trainees were not moving fast enough in the organization and were barriers to each other in the organization. The items concerned with career promise and pay satisfaction loaded on distinctly separate factors, which were labeled A and B in order to underline the different level of conceptualization.

Individual Level Variables and Reactions to the Program

Two types of individual level variables were correlated with the measures of reactions to the program in order to test the hypothesis that there was a relationship between an individual's state in the system and his reactions to the management development program. One kind of variable was based on self-perceptions and included a person's assessment of his career promise and his pay satisfaction.[3] The other kind of individual variables were based on a person's actual position in the system and included his seniority and organizational rank. Table 2 contains the relevant correlations.

There were significant correlations between perceiving one's self as having a positive future and positively evaluating the program, tending not to see the trainees as receiving preferential treatment, tending not to see the program as a barrier, tending not to see the trainees as predominantly Ivy League, and tending to be in favor of more rapid mobility of the trainees. There were also significant correlations between pay satisfaction and positively evaluating the program, tending not to see the trainees as receiving preferential treatment, tending not to see the program as a barrier, and tending not to see the trainees as predominantly Ivy League graduates.

Seniority with the bank and organizational rank were also significantly related to some of the attitudes about the program. Longer service organization members were more likely to evaluate the program negatively, see the trainees as overly privileged, and see the trainees as a clique than newer members of the organization. Higher ranking members of the organization tended to see the trainees as more of a clique and as less of a barrier to their own careers than lower ranking members of the system.

Group Level Variables and Reactions to the Program

Group level analyses of reactions to the program were performed in two ways. The mean scores on each of the attitude scales were computed and compared among the officers, trainees, male employees, and female employees. A multiple discriminant analysis was also performed using these same data.

Table 3 contains the mean values on the

TABLE 3

MEAN FACTOR SCORES ACCORDING TO EMPLOYMENT GROUPS

$\bar{X}$ factor scores	Group[a]			
	Officers ($n = 80$)	Trainees ($n = 22$)	Male employees ($n = 62$)	Female employees ($n = 99$)
Evaluation	−2.10	6.63	−4.36	−2.54
Preferential treatment	14.10	5.59	14.08	12.48
Closed group	4.10	2.46	3.08	1.85
Career barrier	9.35	8.73	12.62	11.96
Ivy League graduates	7.95	5.55	7.82	8.15
Accelerate trainees	.79	5.16	.59	.86
Promising career	2.62	3.42	2.44	2.49
Pay satisfaction	2.49	2.40	1.36	1.89

[a] Overall significance of mean difference was $p = .001$ for all mean factor scores but pay satisfaction (*ns*) by F test with $df = 3,259$.

---

[3] Based on the factor-analytic results, summed scales were developed for each of the scales by giving unit weights to the items defining a particular factor.

attitude scales for each of the groups. There was a significant difference among the groups in the way they evaluated the program. Trainees themselves gave the program the highest evaluation, officers were next, female employees next, and male employees gave the program the lowest evaluation. There was a significant difference among the groups in the degree to which they perceived the trainees as receiving preferred treatment. The officers and the male employees were approximately equally high in seeing the trainees as having and misusing a preferred place in the system; the female employees were slightly lower, and the trainees themselves were substantially lower. There was a significant difference among the groups in the degree to which the trainees were perceived as a closed group. The officers were highest on this perception, the male employees next, the trainees next, and the female employees were lowest. There was a significant difference among the groups in the degree to which they held this attitude. Male employees were most firmly of this opinion, and female employees were only slightly less of the same view. The officers and the trainees themselves were substantially lower. There was a significant difference among the groups in the degree to which they saw the trainees as primarily graduates of Ivy League colleges. Female employees, officers, and male employees were strongly of this opinion while the trainees were less convinced. There was also a significant difference among the groups in terms of the degree of perceived need to accelerate the mobility of trainees as a function of employment group. While there was an overall significant difference among the groups, this difference was primarily a result of the trainees who showed a substantially higher demand for more rapid advancement than any of the other groups, who did not appear to differ substantially from each other.

There was a strong significant relationship between the perception of having a promising future and employment group. The trainees viewed their futures most optimistically, the officers and male employees next, and the female employees least favorably. There were no significant differences among the groups in their pay satisfaction.

All of these group differences provide a measure of empirical validity for the various factors that were structured by factor analysis.

More importantly, however, the group differences point up the fact that there were rather strong group level phenomena that, in part, determined an organization member's reaction to the program. In some sense the group differences were most marked between the trainees and everyone else. Just being a participant in the program led one to evaluate the program more positively, see the members as less privileged, see the program as less of a barrier for others, see fewer of the trainees as Ivy Leaguers, and make more demands for rapid mobility of the trainees. The other groups, however, had their own unique combination of attitudes about the program.

Perhaps the officers had the strongest mixed feelings about the program, for, next to the trainees, they gave the program the highest evaluation. At the same time, however, they held the strongest view that the trainees formed a clique, and they rather strongly felt that the trainees received undeserved preferential treatment. The female employees had a different combination of attitudes, a predominant theme of which was undoubtedly romantic. Their evaluation of the program was almost as high as that of the officers, and they saw the trainees as less of a closed group than the trainees themselves did. At the same time, however, they saw the trainees as abusing their status and as blocking women's careers. For male employees, the emerging picture of the program was almost entirely dismal. They gave the program a low evaluation, saw the trainees as abusing their preferred status, and saw the program as a barrier to their own careers. In a very real sense the program pitted the trainee against the male employee and the attitude results reflected this fact.

To further explore the role of group effects in determining reactions to the program, a multiple discriminant analysis was performed using scores from the eight scales to differentiate among the four groups. Multiple discriminant analysis is a statistical tool that determines how a set of variables should be weighted to discriminate maximally among a set of a priori groups. A discriminant function is a set of weights that defines a vector. In the present analysis, normalized vectors are re-

TABLE 4

MULTIPLE DISCRIMINANT ANALYSIS OF THE ATTITUDE DIMENSIONS FOR THE FOUR EMPLOYMENT GROUPS

Attitude scales	Discriminant vectors		
	1	2	3
Evaluation	.08	−.02	.16
Preferential treatment	−.09	.22	.09
Closed group	.34	.37	.08
Career barrier	−.33	−.62	−.59
Ivy League graduate	−.15	.13	.32
Accelerate trainees	.71	−.49	.15
Promising future	.42	.32	−.67
Pay satisfaction	−.22	.24	.15
Ratio of between to within variance	77**	27**	21*

* $p = .01$.
** $p = .001$.

ported. These vectors are standardized; they do not reflect differences in variation in the variables used and, therefore, are analogous to beta weights in multiple regression.

Table 4 presents the normalized vectors for differentiating among the employment groups. The first vector is characterized by substantial positive weights on accelerating the progress of trainees by people who saw themselves as having a promising future with the bank. The second vector is characterized by large negative weights on the program as a career barrier and on accelerating the progress of trainees. The third vector has substantial negative weights on the program as a career barrier by people who tend not to see themselves as having a promising future with the bank.

The means of the three functions for each of the employment groups are shown in Table 5. Vector 1 serves primarily to discriminate the trainees from the other three groups and especially from the female employees. Trainees would be readily characterized as wanting to see themselves move more rapidly up the hierarchy and having a promising future, while the female employees would have almost the opposite views. The second vector seems primarily to emphasize the differences between the officers and the trainees. One can understand the officer discrimination by noting that trainees do not pose a threat to the current officers' careers, providing that the trainees are not accelerated too rapidly. Finally Vector 3 serves mainly to discriminate the male employees from the other groups. This vector emphasized the reactions of those who saw the program as a barrier to their own careers and yet saw themselves as having a promising future with the organization.

The multivariate analysis called attention to the role of group level effects in a way that univariate analysis alone did not. Although the global evaluation scale showed substantial differences among the four groups when the analysis was univariate, it had consistently low weights on all of the discriminant functions. Perhaps this is a way of showing that what was at stake for the groups was not whether the program was good for the bank. Starting it positively, one might make a case for group self-interest as well as individual self-interest.

Combined Impact of Group and Individual Factors

In addition to considering the separate effects of individual and group effects on a person's reaction to the program, there was also the possibility of finding that there would be an interaction of individual and group effects. In other words, a person with certain individual qualities might react differently to the program if he was a member of one group than if he was a member of another group. It turned out that for perception of career promise, at the individual level, and for the officer, trainee, and male employee groups there were statistically significant interactions observed (see Table 6). Trainees who perceived themselves to have promising careers were less likely to see the program as a closed group in which the members associated mainly with each other than were either the officers or the

TABLE 5

MEANS OF THE FOUR EMPLOYMENT GROUPS IN THE DISCRIMINANT SPACE

Groups	Discriminant axes		
	1	2	3
Officers	3.81	−9.02	1.24
Trainees	−2.12	−5.64	1.55
Male employees	4.72	−7.32	3.62
Female employees	5.82	−6.20	1.10

male employees who perceived themselves to have a promising career. Trainees who perceived themselves to have a promising career were less likely to see the program as a career barrier or the trainees as predominantly Ivy Leaguers than were the male employees. These interactions underline in yet another way the degree to which the program was a psychologically different thing for the aspiring male employee who perceived that he was successful than for the trainee who perceived that he had been successful.

TABLE 6

SIGNIFICANTLY DIFFERENT CORRELATIONS BETWEEN PERCEPTION OF A PROMISING FUTURE AND REACTIONS TO THE PROGRAM FOR DIFFERENT EMPLOYMENT GROUPS

Reactions to the program	Group				
	Officers ($n = 80$)		Trainees ($n = 22$)		Male employees ($n = 62$)
Closed group	.13	**	−.59	**	.30
Career barrier	−.18		−.48	*	.07
Ivy League graduates	−.01		−.47	*	.06

* Difference between correlations significant at $p < .05$ (two-tailed test).
** Difference between correlations significant at $p < .01$ (two-tailed test).

DISCUSSION

There can be little doubt from these results that reactions to the management development program were a function of not only individual but also of group level variables. Individual variables such as perceived career promise, pay satisfaction, seniority, and organizational rank were related to two or more different reactions to the program. A person's reaction to the program also turned in a large measure on which employment group he was a member of. The various employment groups turned out to have their own unique patterns of attitude clusters about the program, which in part also reflected the interaction of group and individual level variables.

The emphasis on intergroup dynamics could overshadow the existence of intragroup conflict among members of the program. But the sixth factor among the attitude scales drew attention to this phenomena in an unmistakeable way. Only trainees were concerned that they were not moving fast enough in the organization. For those so concerned, it was often other trainees who formed their career barriers. This point received an interesting independent replication in a report on the program prepared by one of the program administrators prior to this study and made available to the present author after his own report had been prepared. In it the program administrator concluded:

> It is my opinion that our excessive turnover within the middle group was caused . . . by a frustration of opportunity within the minds of the trainees comprising the middle section. The surface reasons for leaving were varied. But beneath the surface, whether realized or unrealized, the middle group of trainees looked upon the older group of trainees as direct competitors for similar positions. By virtue of their presence, the older group, as long as they continued to demonstrate their ability within a particular area, pre-empted to a degree the movement of the middle group towards responsibilities and positions they were seeking.

Behavioral Explanations for the Attitude Effects

By virtue of being members of the same organization, the trainees and the other groups had regular opportunities to interact with each other both within and between groups. The process of gathering data about the program and feeding it back to the organization provided an opportunity to observe how the various organization members behaved toward one another and to ask the organization members what they perceived in each other's behavior. These various behaviors are offered as some mediating variables to explain why the attitudes about the program came to be what they were. There is, of course, no claim of exhaustiveness for these behaviors. No explanatory behavior is reported unless it was observed or reported more than once.

Mechanism 1: The meeting of the "right types" and the "new types" result in contrasting styles of dealing with aggressiveness without having the capacity to discuss openly the consequences of the different styles. At the time when Argyris (1954) studied this organization, he reported that the modal employee and manager tended to avoid being aggressive himself and to avoid being in situations where aggressiveness would be present. When the present study was conducted, the organization was no longer populated as exclusively with nonaggressive types. Most of the trainees and many of the younger officers and employees were much more comfortable with aggression in themselves and in others than the old right types. But when the two types interacted, the contrasting styles tended to block valid communication.

In a group interview, the following exchange took place:

> Older "right type": "Do people find a frustration of opportunities because of the training program?" (He was interrupted in midsentence by a trainee who said, "Absolutely not!")
>
> Trainee: "I don't think so! I don't think so! We have always had internally excellent reception of trainees, cooperation from people. . . . They've accepted them readily, happily. There has been no frustration."
>
> Right type: "Well, I can't quite agree with that. I don't think it's. . ."
>
> Right type: "In *our* particular department . . ." (Interrupted by trainee).
>
> Trainee: "Oh, in your area. I used to get quite a bit of flack from some of the lads in the back that were not college grads and thought that their potential as far as the bank was concerned was severely limited by trainees. . ."

This kind of pattern could often be observed when the contrasting styles encountered one another on a subject about which both had strong and differing opinions. The pattern could be characterized by the aggressive style escalating the forcefulness of the discussion, which would be responded to by withdrawal from the nonaggressive type. This would frustrate the aggressive type who would then escalate his aggression, but that would only frighten the nonaggressive type more, and he, in turn, would withdraw further.

Mechanism 2: By means of a self-fulfilling prophecy the interpersonal pattern was almost institutionalized. Although the trainees did not talk with their antagonists about the problems, they did talk with each other. New trainees were warned to expect trouble getting information in certain areas and were advised to be more aggressive in order to cope with the resistance.

> Graduate trainee: "I think possibly there may be a few people who are gunning for the trainees as a group. There's a little resentment that these are young guys that are coming along quickly; they're going to be officers quickly. . . . So, I think maybe it's a concept they may be gunning for, but as individuals, it's going to be up to you whether they are gunning or not. . ."
>
> Trainee: "You've got to press these people for the information. You've got to drag it out of them because, after all, they know you're going to become an officer . . . a lot quicker than they will ever be . . . If you go into an area and you simply keep on pressing these people they're going to give you the information as long as you show the initiative. . ."
>
> Trainee: "There are people that will give you as much as they think you need, but you gotta dig like hell to get any more out of them."
>
> Trainee big brother (telling how he advises his trainee): "Now look, X, this department is going to be a difficult department for you to get into and this is what's expected of you. You're a real bright guy; you graduated from Yale; I know you're a fighter. You're dynamic and aggressive. Show them that you are. Go in there and fight. . ."

One could speculate further that by encouraging the newer trainees to express their aggression in the departments the big brothers were (intentionally or unintentionally) discharging some of the *intragroup* conflict by channeling it into *intergroup* conflict.

Mechanism 3: Both the most negative and the most empathic views of the program tended to be shown without the presence of both parties to the conflict. One place where a person was free to express his views of the program was on the anonymous questionnaire dealing specifically with the program. Some of the most extreme views included the following:

> If these are some of the comments made by employees, they certainly are either very shortsighted or 'sour grapes' comments. . . . Questions 5, 19, and 24, although valid objections perhaps, in some cases, are not directed at 'trainees' but individuals [written by a trainee].
>
> Questions 5, 19, and 24 express my complete disgust with the program. The program is basically a good means to build the management team. However, it is ruined by the attitude and behavior of the trainees [written by a nontrainee officer in his mid-30s].

The same instrument, however, produced some reactions that showed that both sides could see that there were two sides to the problem.

> I suspect there are definite inequities in terms of non-trainee-employee evaluations and opportunities for these employees to seek out and complete the program. . . . Perhaps if more were aware of the op-

> portunities and requirements of the program, and if the college education requirement was waived if other aptitudes offset it, more would come forward, be accepted and boost morale, and ease some ill feeling within the bank [written by a trainee].
>
> Not being a college graduate and not qualified to enter the Training Program and progress more rapidly in the bank, my instinct for survival pledges me against the program. Deep down inside, I know it means for a better bank [written by a male nontrainee in his mid-20s].

Mechanism 4: Most of the initial recruitment, selection, and evaluation of trainees was done either by former trainees or by the top most members of management, resulting in other members of management feeling as though they had almost no influence on the program. In addition, there was almost no source of public information or open discussion about how the program operated so that people could check their perceptions against some measure of reality. As the issue was discussed time and again, it became apparent how the operation of the preceding mechanisms may have influenced the decisions that brought about this structural pattern.

Action Implementation

As was mentioned above, considerable attention was given to the results of this study by members of management at the bank. After the feedback meetings had been completed, a task force to consider changes in the training program was formed. This group was chaired by the Vice-President for Personnel and consisted of the Program Administrators, some current trainees, some younger officers who were not graduates of the program, and some older officers. One of the first things that this group discovered was that they were lacking in information about the program. As a result, the program administrators prepared an information memorandum providing such facts as: (*a*) which colleges were visited for recruiting, (*b*) other methods for making the program known to potential applicants outside the bank, (*c*) the actual proportion of trainees from Ivy League colleges, (*d*) the number of trainees who entered from inside the bank, (*e*) the number of trainees who had been asked to resign from the program, and (*f*) the percentage of retention of trainees from Ivy League versus other colleges. This memorandum was sent to all personnel in the bank, and it was reported to this writer that the reaction to it was very positive.

In addition, the task force also made a number of recommendations for structural changes in the program. First, they recommended that more people than the Program Administrators should be involved in the recruiting interviews and that these interviewers should be assigned by the Personnel Department. Second, they recommended that the pool of people from whom big brothers were assigned to trainees should be broadened to include more people than just former trainees and that these counselors should be assigned by the Personnel Department. Finally, they recommended that each trainee should be assigned a research paper on some general topic in banking (e.g., "Banking by Computer—1980"). The opportunity to do a project like this, and thereby be counseled by a senior officer, should be opened to anyone in the bank. At the time of this writing, all of these recommendations had received the informal approval of senior management, and the second one was already being implemented.

REFERENCES

ARGYRIS, C. *Organization of a bank.* New Haven: Labor and Management Center, Yale University, 1954.

ARGYRIS, C. *Understanding organizational behavior.* Homewood, Illinois: Dorsey, 1960.

DUNNETTE, M. D. *Personnel selection and placement.* Belmont, California: Wadsworth, 1966.

EMERY, F. E., & TRIST, E. L. The causal texture of organizational environments. *Human Relations*, 1965, **18**, 21–32.

ETZIONI, A. *Modern organizations.* Englewood Cliffs, N. J.: Prentice-Hall, 1964.

GLICKMAN, A. S., HAHN, C. P., FLEISHMAN, E. A., & BAXTER, B. *Top management development and succession.* New York: Macmillan, 1968.

KATZ, D., & KAHN, R. L. *The social psychology of organizations.* New York: Wiley, 1966.

RICE, A. K. Individual, group, and intergroup processes. *Human Relations* 1969, **22**, 565–584.

SCHEIN, E. H. *Organizational psychology.* Englewood Cliffs, N. J.: Prentice-Hall, 1965.

SIMON, H. A. *Administrative behavior.* (2nd ed.) New York: The Free Press, 1957.

(Received April 16, 1970)

Vol. 53, No. 4, 312–316

MOTIVATION IN MANAGERIAL LEVELS: RELATIONSHIP OF NEED SATISFACTION TO JOB PERFORMANCE[1]

JOHN W. SLOCUM, JR.[2]

Department of Management Science and Organizational Behavior,
Pennsylvania State University

This study compares the need satisfaction of first-line supervisors with top and middle managers and relates need satisfaction to job performance. A need satisfaction questionnaire was used to collect the data from managers in a steel mill. The findings indicate support for Porter and Lawler's model relating need satisfaction to performance, but only partially support the hypothesis that satisfaction of higher order needs is more closely related to top managers performance than satisfaction of higher order needs for lower managerial personnel.

During the past decade, the content of job satisfaction studies of different organizational levels has shifted from comparisons of operative personnel versus supervisory personnel to comparisons among different levels of management within an organization. The major focus of these studies seems to be whether satisfaction increases or decreases with each higher level of management.

In studies reporting on the differences between managerial levels, Porter (1961) and Rosen (1961a, 1961b) have offered contradictory findings. Rosen found that top managers in the plant and middle managers do not differ significantly from each other in satisfaction, but these managers were significantly more satisfied than first-level supervisors. In Porter's study, two levels of management were studied using a questionnaire designed to tap satisfaction of Maslow's (1943) hierarchical classification of needs. The results of this study indicate that management personnel differ in their need satisfaction, although the patterns of need satisfaction were similar.

In a much larger study of managers, Porter's (1962, 1964) results supported his earlier findings. That is, need satisfaction increased with each hierarchical level of management for the higher order needs (esteem, autonomy, and self-actualization), while the lower order needs (security and social needs) were almost equal across all managerial levels. Similar results have been reported by the Opinion Research Corporation (1962), and Haire, Ghiselli, and Porter (1963). Therefore, from the evidence reported to date, it can be stated that job satisfaction increases at each higher level of management and the patterns of need satisfaction are similar across different organizational levels within the managerial rank.

Relationship between Satisfaction and Performance

In a recent study reported by Lawler and Porter (1967) and also in their book (Porter & Lawler, 1968), these researchers posit a theoretical model that indicates that performance leads to rewards, which may in turn lead to satisfaction. They argue that satisfaction, rather than causing performance, is caused by it. The model extended by these researchers distinguished between two types of rewards. Extrinsic rewards, such as pay, promotion, and security are likely to be imperfectly related to performance. These rewards are often primarily associated with satisfaction of the lower order needs of safety and security discussed by Maslow (1943).

Lawler and Porter (1967) posit the opposite case for intrinsic rewards. They comment that these rewards, for example, the feeling of having accomplished something worthwhile, are more directly related to performance. The

---

[1] This research was financed by a College of Business Administration Faculty Research Grant from the Pennsylvania State University and a research grant from the Standard Steel Division of Baldwin-Lima Hamilton Corporation. The researcher wishes to acknowledge his indebtedness to William Gray, Vice-President of the Standard Steel Division, for his time and cooperation in this study.

[2] Requests for reprints should be sent to John W. Slocum, Jr., Pennsylvania State University, Department of Management Science and Organizational Behavior, 223 Boucke Building, University Park, Pennsylvania 16802.

model also indicates that neither intrinsic rewards nor extrinsic rewards are directly related to job satisfaction, since the relationship is moderated by the amount of expected equitable rewards. Expected equitable rewards refer to the amount of rewards that an individual feels he should receive as a result of his job performance (Porter, 1961; Porter & Lawler, 1965). Thus, an individual's satisfaction is a function both of the magnitude and the frequency of rewards he receives as well as his expected equitable rewards.

The model predicts a low but positive relationship between job satisfaction and performance because of the imperfect relationship between performance, rewards, and satisfaction. However, very few studies could be found that systematically consider satisfaction of needs in terms of managerial performance, and studies relating satisfaction of needs at different organizational levels to performance have not yet been reported in the literature.

A study by Lawler and Porter (1967) correlated satisfaction of managers' needs with performance. All correlations were significant but again low, ranging from .16 to .30. The researchers conclude that satisfaction of higher order needs is more closely related to performance than satisfaction of lower order needs, but they fail to test for statistical differences between correlations. The correlations for the lower order security and social needs were .21 and .23, respectively, while the correlations for the higher order needs of esteem, autonomy, and self-actualization were .24, .18, and .30, respectively. The differences between the correlations do not appear to be great enough to warrant their conclusion that higher order needs are more closely related to performance than lower order needs. Porter and Lawler (1968) provide additional data for support of their model. Higher order needs accounted for more significant relationships at the .01 level than lower order needs. However, they do not report correlations between these needs and performance. Therefore, no evaluations can be made regarding the relationship between need satisfaction and performance.

The purpose of this article is to reexamine the earlier findings relating need satisfaction to managerial levels and also examine the model posited by Porter and Lawler in terms of performance. This article considers one specific prediction: Satisfaction of higher order needs is more closely related to top management performance than satisfaction of higher order needs for lower management.

TABLE 1

SUMMARY OF PROBABILITIES OF THE TESTS FOR CONVERGENT AND DISCRIMINANT VALIDITY

Validity	Peer-superior rating					
	1	2	3	4	5	6
Convergent validity	.01	.01	.05	.05	.01	.01
Discriminant validity tests						
Test 1	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
Test 2	.01	.05	.01	.01	.01	.01
Test 3 (Overall)[a]	< .01					

[a] $W = .80$, $\chi^2 = 43.87$.

METHOD

Data were collected from 87 first-line supervisors and 123 top- and middle-level managers employed by a steel plant in central Pennsylvania. Two kinds of data were collected for each manager by the researcher and his assistant. The job performance scale used for this study was a multitrait scale developed by the director of training. The traits that the manager was rated on were: (*a*) technical knowledge, (*b*) functional knowledge, (*c*) drive/aggressiveness, (*d*) reliability, (*e*) cooperation, and (*f*) organizing ability. For each trait, the rater chose one of the five responses—superior, fully competent, shows improvement, getting by, and poor. The raters used were *S*'s superior and peers. In the choice of the raters, the constraint was that they must have been familiar with the aspects of the individual's performance that they were asked to rate. To insure the choice of knowledgeable raters, *S* was asked to name three peers who knew his performance best. The choice of which peer rated *S* was done randomly.

The performance measures were tested for convergent and discriminant validity according to the criteria posited by Campbell and Fiske (1959). The purpose and interpretation of the tests for discriminant validity as suggested by Campbell and Fish are briefly described below.

Test 1. The correlations among two measures of a single trait should be greater than the correlations between any one rating of that trait and any other trait rated by the other rater. Failure to meet this requirement was due to the high intercorrelations. According to Lawler (1967), this is a very stringent requirement that is seldom met by behavioral trait data.

Test 2. The correlation between two ratings of a single trait should be greater than the correlation between any one rating of that trait and any other trait

TABLE 2

AVERAGE NEED SATISFACTION SCORES FOR MANAGERS AT THREE DIFFERENT LEVELS

Need category and item	First-line managers	Middle-top managers
1. Security need		
Security in job	2.12	1.19
2. Social needs		
Opportunity to help people		
Opportunity for friendships	1.39	1.13
3. Esteem needs		
Feeling of self-esteem	1.54	.90*
Prestige inside the company	1.52	.56**
Prestige outside the company	1.44	.81*
4. Autonomy needs		
Opportunity for independent thought and action	1.39	1.19
Authority in position	1.40	.68**
Opportunity to participate in goal-setting	1.53	1.17*
5. Self-actualization needs		
Opportunity for growth and development	1.35	.97**
Feeling of self-fulfillment	2.11	.62**
Feeling of accomplishment	1.38	.41**

* $p = .05$ (two-tailed t test).
** $p = .01$ (two-tailed t test).

rated by the same rater. This test was supported by the validity coefficients being greater than the coefficients in the monorater-multitrait triangles at the .05 level.

Test 3. There should be the same pattern of trait interrelationships in all the heterotrait triangles of both the monorater and heterorater blocks. The data supported this test as indicated by the significant W in Table 1.

From Table 1, it can be seen that all six of the peer-superior coefficients (correlations ranged from .56 to .86) met the test for convergent validity at the .05 level or better. Computing Kendall's W coefficient of concordance (Siegel, 1956) to determine the similarity of patterns between raters, a coefficient of .80 was obtained. Thus, the raters were essentially in agreement on the ranking of *S*s under study. It should be emphasized that the significant W does not mean that the rankings were correct; it implies that the raters all agreed on the patterns.

The data for both groups of employees' need satisfaction was obtained by means of the Porter (1961) questionnaire. The questionnaire contained 12 items that were predicated upon Maslow's theory of motivation. All *S*s were provided with the following form:

> On the following pages will be listed several characteristics or qualities connected with your own job. For each characteristic, you will be asked to give two ratings:
>
> a. How much of the characteristic do you think should be connected with your job?
> b. How much of the characteristic is there now connected with your job?
>
> Each rating will be on a seven-point scale, which will look like this:
>
> (minimum) 1 2 3 4 5 6 7 (maximum)
>
> You are to circle the number on the scale that represents the amount of the characteristic being rated. Low numbers represent low or minimum amounts, and high numbers represent high or maximum amounts.

The answer to the first question was taken to measure the employee's expected equitable level of rewards. The second question was assumed to be a measure of need fulfillment. The difference between the first and second answer determined the amount of perceived need deficiency of each item. Thus, the larger the difference between "should be" and "is now," the greater the dissatisfaction.

The items, presented in random order, were preclassified into the five types of needs described by Porter (1961): security, social, esteem, autonomy, and self-actualization.

RESULTS AND DISCUSSION

The data collected from the sample supports the previous work done by Porter (1962, 1964) and Rosen (1961a, 1961b). The data in Table 2 indicate that satisfaction generally increased each level of management. Except for one need item in the first-line supervisors' self-actualiza-

TABLE 3

CORRELATIONS BETWEEN PERFORMANCE AND SATISFACTION ON ITEM SCORES IN EACH NEED CATEGORY FOR MANAGERS AT THREE DIFFERENT LEVELS

Need category and item	First-line managers' performance	Middle-top managers' performance
1. Security need		
Security in job	.13	.11
2. Social needs		
Opportunity to help people	.16	.20
Opportunity to develop friendships	.18	.18
3. Esteem needs		
Feeling of self-esteem	.17	.20
Prestige inside the company	.15	.29
Prestige outside the company	.19	.29
4. Autonomy needs		
Opportunity for independent thought and action	.22	.33
Authority in position	.15	.38*
Opportunity to participate in goal-setting	.20	.42*
5. Self-actualization needs		
Opportunity for growth and development	.31	.39
Feeling of self-fulfillment	.22	.36
Feeling of accomplishment	.18	.42*

Note.—A correlation of .18 was required to be significant ($p < .05$) for the foremen sample. A correlation of .17 was required to be significant ($p < .025$) for the middle-top manager sample.

* $p = .05$ (two-sided test between *r*s for first-line vs. middle-top managers).

tion category, the need satisfaction profiles are similar. For the security and social needs, satisfaction was roughly equal for both levels of management. The higher levels of management, on the whole, reported greater degrees of need satisfaction in their jobs than did lower-level managers. In no instances did higher levels of management express greater dissatisfaction than did the lower level. It should be indicated that for the lower order need of security, first-line and middle-top managers were not getting less fulfillment, but that there was a larger difference between their expectations and their fulfillment than was found at the higher order need level.

The Lawler and Porter (1967) model posits that an individual's degree of higher order need satisfaction is related to his performance. The current research supports this general prediction. As predicted, the higher order needs of autonomy and self-actualization were more closely associated ($p < .05$) with performance than the lower order needs for both first-line supervisors and middle-top level managers. This finding did not carry over for the esteem need. The data in Table 3 indicate these relationships.

The highest average correlation was found between performance and self-actualization for both groups of managers. In testing for significant differences between the correlations in Table 3, only three differences were statistically significant at the .05 level. The correlations for the middle-top level managers autonomy needs were statistically different in two need items, and one self-actualization need item (feeling of accomplishment) was statistically different between the two groups in that need category. Thus, the prediction that satisfaction of higher order needs is more closely related to middle-top managerial performance than for lower managerial performance is supported by these data. Although only three differences (see Table 3) were significant, differences for items in the esteem, autonomy, and self-actualization categories were in the predicted direction.

The data collected from the sample also indicate that job satisfaction (the sum of the difference score on all 12 items) correlated significantly with performance for both groups of managers. Although the correlations were not large ($r = .26$, $p < .01$, for first-line supervision; $r = .42$, $p < .001$ for middle-top managers) it is substantially different from that originally reported by Lawler and Porter (1967; $r = .32$, $p < .01$) in their study of middle and lower-middle managers and that reported by Vroom (1964).

Overall, the data from the present study are in general agreement with the predictions forwarded by Lawler and Porter's (1967) model concerning the relationship between higher order need satisfaction and performance. Similarly, the relationship between need satisfaction and managerial level has been supported. Patterns of need satisfaction were similar across organizational levels, and the higher order needs were more satisfied for the middle-top level managers than for the first-line supervisors. While statistical differences between need satisfaction and job performance were found in only three need items, these differences were found in the higher order needs. This indicates that for these needs, middle-top level managers' higher order needs are more closely related to their job performance than the satisfaction of higher order needs for first-line supervisors.

The data also indicate that satisfaction of lower order needs are not correlated with performance for first-line supervisors. While statistically significant correlations were noted for middle-top level managers in each need category except for the security needs, the satisfaction of higher order needs are more closely associated with performance ($p < .05$) than the satisfaction of lower-order needs. Since there has been relatively little evidence concerning Maslow's (1943) need theory and managerial performance (Hunt & Hill, 1969), future studies using the same measures should be undertaken to examine the relationships between organizational level and performance.

REFERENCES

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 1959, 81–105.

Haire, M., Ghiselli, E., & Porter, L. W. Cultural patterns in the role of the manager. *Industrial Relations*, 1963, 2, 95–117.

Hunt, J., & Hill, J. The new look in motivation theory for organizational research. *Human Organization* 1969, 28, 100–109.

Lawler, E. Multitrait-multirater approach to measur ing job performance. *Journal of Applied Psychology*, 1967, 51, 356–381.

LAWLER, E. & PORTER, L. W. The effect of performance on job satisfaction. *Industrial Relations*, 1967, 7, 20–28.

MASLOW, A. A theory of human motivation. *Psychological Review*, 1943, 50, 370–396.

OPINION RESEARCH CORPORATION. *Motivating managers*. Princeton, N. J.: Author, 1962.

PORTER, L. W. A study of perceived job satisfaction in bottom and middle management jobs. *Journal of Applied Psychology*, 1961, 45, 1–10.

PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, 46, 375–384.

PORTER, L. W. *Organizational patterns of managerial job attitudes*. New York: American Foundation for Management Research, 1964.

PORTER, L. W., & LAWLER, E. E., III. Properties of organization structure in relation to job attitudes and job behavior. *Psychological Bulletin*, 1965, 64, 23–51.

PORTER, L. W., & LAWLER, E. E., III. *Managerial attitudes and performance*. Homewood: Irwin-Dorsey, 1968.

ROSEN, H. Desirable attributes of work: Four levels of management describe their job environment. *Journal of Applied Psychology*, 1961, 45, 156–160. (a)

ROSEN, H. Managerial role interaction: A study of three managerial levels. *Journal of Applied Psychology*, 1961, 45, 30–34. (b)

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1966.

VROOM, V. *Work and motivation*. New York: Wiley, 1964.

(Received April 15, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 317–323

A STUDY OF ENTREPRENEURIAL MOTIVATION: COMPARISON OF FAST- AND SLOW-PROGRESSING SMALL-SCALE INDUSTRIAL ENTREPRENEURS IN PUNJAB, INDIA[1]

P. S. HUNDAL[2]

Punjab University, Chandigarh, India

The purpose of this study was to determine the level of achievement motivation and other motivational factors associated with the differential rate of industrial growth of the fast- and slow-progressing entrepreneurs. One hundred and eighty-four small scale industrial entrepreneurs were administered tests of achievement motivation, attitude toward money, and attitude toward labor unions, and interviewed to assess job satisfaction, job prestige, value patterns, aspirations, and interests. The analyses show that aspirations, achievement motivation, and investing tendency are associated with a fast rate of industrial growth, whereas, hoarding tendency and optimistic tendency seem to be related to the slow rate of industrial growth.

The economic survey of small-scale industrial entrepreneurs of Punjab, India (Rangnekar, 1966), located two groups of sharply contrasting small-scale industrial entrepreneurs in Ludhiana, India, a renowned center of small-scale industries in the country. One of these groups was engaged in metal-using industries (metal group) and the other in the hosiery and textile industries (hosiery group). The exponential growth rate for the two groups, over a period of 10 yr., 1951–1961, was 32% and 10%, respectively. Rangnekar also reported that the difference in the rate of growth of the two groups could not be explained satisfactorily in terms of economic factors. He hypothesized that there might be some social and psychological variables influencing the rapid growth of the entrepreneurs engaged in metal-using industries.

The role of social factors in the development of economic growth has, indeed, been stressed much earlier (e.g., Hagen, 1961; Rostow, 1961; Weber, 1930). More recently, McClelland (1961, 1963) sought to show that the essential ingredient of entrepreneurship is a social value called achievement motivation (n Ach). This he describes as a desire for excellence, not so much for the sake of social recognition, but to attain an inner feeling of personal accomplishment. He also equates it with inner concern for achievement, a disposition to engage in activities in which doing well in competetion with others is important.

Studies by McClelland (1961), Bradburn (1960), Bradburn and Berlew (1961), Cortes (1961), and many other investigators have presented a wide range of evidence indicating a relationship between level of achievement motivation and national economic development. The relationship, however, was of low magnitude. Nations were used as units of analysis in these studies. While this type of data is of great importance, evidence in which the individual is the unit of analysis is also needed.

The present study is a step in this direction. Its main purpose is to determine the level of achievement motivation in a group of fast- and slow-progressing small-scale industrial entrepreneurs of Punjab, India, and also to find other motivational factors that characterize the two contrasting categories of entrepreneurs.

METHOD

Subjects

One hundred and eighty-four small-scale industrial entrepreneurs comprised the sample. They were drawn from the industrial population of Ludhiana, India.

In drawing this sample, all 2,308 industrial units (excluding one-man units) were enumerated and classified into 10 categories. Of these, two major categories were chosen, the textile industries (inclusive of hosiery) accounting for 1,390 and the metal-using industry accounting for 571. The two categories cover 85% of all units. Once this population was fixed, 200 sample units, 100 from each category, were taken at random for the study. Due to one or the other reason, however, only 184 units could be included in the study.

---

[1] This project was supported by the Asia Foundation, New Delhi, India.

[2] Requests for reprints should be sent to P. S. Hundal, Department of Psychology, Punjab University, Chandigarh, India.

TABLE 1

ITEM 1: "IN THE NEXT FIVE YEARS I AM GOING TO . . ."

Point score	Typical response
7	Manufacture top-class goods; establish a big factory, a big mill, or big business; have more than 1,000 workers in my factory; extend my work by having a maximum number of workers.
6	Extend my factory or business; improve the quality of goods manufactured in my factory; increase the ouput; get some new machinery; earn more by making my business flourish.
5	Be a good business man (good in dealings); be a municipal commissioner; be a rich man (field of activity not mentioned).
4	Do my present business; have a better life; extend my business if government will give a loan (element of doubt present).
3	Increase production if circumstances will permit; join some other friend in his business (somewhat like an inactive partner).
2	Be in the same position; earn enough to meet my daily needs.
1	Leave this business.

The age of the sampled entrepreneurs ranged between 25 and 45 yr. with a mean and standard deviation of 39 yr. and 10 yr., respectively. With a few exceptions, almost all of them belonged to the middle strata of the current caste pyramid of Punjab. Twenty-nine percent of them were immigrants from West Punjab and about the same number from other districts of East Punjab. The forefathers about two generations back of the greater majority of the *S*s were agriculturists.

Test Variables

Achievement motivation test. The "achievement motivation" (AM) was measured by a modified version of Rogers & Neill's (1966) sentence completion test on AM. It consists of 12 items designed to elicit responses reflecting the level of AM of *S*s. The responses were rated on a 7-point scale ranging from 1 to 7 following Rotter and Willerman (1947) method, the higher scores denoting greater AM. The total score had a numerical range of 12–84.

In selecting and weighing responses for the different categories, McClelland's (1961) description of a person with high n Ach was kept in view. Consequently, the responses showing persistent and energetic striving for business and industry were given the highest score of seven. The lowest score of one was assigned to responses showing a very low degree of striving for progress and persistence in industry or business.

The scaling of typical responses for each item was done by E in consultation with two psychologists (one was a clinical psychologist). A scoring booklet with typical responses for each test item was thus developed. The actual scoring examples for Item 1 are presented in Table 1.

The test-retest reliability coefficient of the test measures with a time gap of about 3 mo. for 35 *S*s, drawn from the sample, was .84.

The item homogeneity or item validity coefficients, following Adams (1960), were obtained by determining phi coefficients between the single item and the total score on the test. The phi coefficients were computed for 27% high and 27% low scores. With the exception of Items 3 and 9, all the phi values were significant at the .01 or .05 level. It was inferred, therefore, that the test items differentiated well between persons with high and low degrees of AM. The discriminating power of Items 3 and 9 was rather low, but they were not excluded from the test.

Further, the scores obtained for 60 *S*s by the sentence completion test of AM were compared with their n Ach scores obtained by a modified version of McClelland's Thematic Apperception Test (1961 TAT).[3] The rank order correlation between the two sets of scores was as high as .61. The fact suggests that there is much in common between the measurements made by the two tests.

Attitude toward money. The test consisted of 30 incomplete sentences with money as a key word. The *S*s were required to complete in a meaningful way the sentence as quickly as possible. The responses on the test were scored in terms of (*a*) tendency to invest money (Inv T); (*b*) tendency to hoard money (Hoa T); (*c*) tendency to spend money on social work (Soc W); (*d*) tendency to spend money on personal and family comforts (P Comf); (*e*) responses showing indecisiveness and lack of confidence, conflicting attitude (Conft A); (*f*) responses showing faith in future, optimism (Opt T). In all cases the responses were categorized at their face value and no attempt was made to infer any underlying unconscious trends.

The responses for the total test of 30 items were tallied under the above categories of responses for each *S* separately. A small number of responses that could not be grouped under any of these categories were ignored. The total number of responses made by an individual in a particular category comprised his score for that variable.

As a check on the scoring procedure, 30 test blanks were rescored by an independent judge. There was 94% agreement on the items scored by the writer and the independent rater.

The consistency of responses was assessed by retesting 35 *S*s after a period of about 3 mo. The correlation coefficient between the initial and the final scores for different categories of responses ranged between .37 and .82. The retest coefficients were very low for the categories Soc W and P Comf. No reasons are offered for their instability.

The validity of the categorized responses was checked following a method suggested by Rohde (1946). Three entrepreneurs selected from the metal group were asked to select from the participants some persons whom they knew intimately. Thirty persons were drawn. They were then asked to rate each of these persons on a 10-point scale to determine their standing on the test variables of Inv T, Hoa T, Soc W, P Comf, Conft A, and Opt T. The final scores for each *S* on each of the rated variables were found by averaging the independent rating of the judges.

---

[3] Modification done by the staff of SIET Institute, Hyderabad

The set of scores thus obtained for these 30 *S*s was then correlated with their respective scores on the sentence completion with money. These correlations ranged between .45 and .81, the lower values being for Conft A and Opt T. The low correlations can be attributed to error of judgment by the rater or their failure to grasp the complex concepts of conflicting attitude and optimistic outlook. It may also be due to some error in categorizing the responses. However, all the correlations were positive and in the expected direction and it seems safe to infer that the sentence completion test with money is a reasonably satisfactory instrument for assessing the strength of the various aforementioned trends.

Attitudes toward labor unions. Attitude toward labor unions (Att L) was measured by a test following Thurstone's (1929) technique of equal-appearing intervals. It had 17 statements that were selected on the basis of the responses of 50 judges. The *Q* values for the selected items ranged between .19 and 1.74 with a mean of 1.21. The range of weight scores was between 0 and 10, the lower score denoting a favorable attitude, and the higher vice versa.

Interview schedule. The data for job prestige, job satisfaction, value patterns, and aspirations were obtained by interviewing *S*s and asking them some set questions designed to elicit responses in the specified area. The *S*s were completely free to state their responses. They could thus express their true reactions.

Their responses were then rated on a 5-point scale in consultation with some teachers and students of psychology who had no prior information about the growth rate of *S*s.

In case of job satisfaction, higher scores were given to responses that stated liking for the present job and a tendency to stick to it.

Regarding job prestige, *S*s were required to rate on a 5-point scale their jobs and the activities related to their jobs. The total score for each *S* was found by adding the rating scores.

The problem of quantifying the value pattern was rather complicated, especially because of disagreement about the hierarchy of the expressed values. It was, however, decided to assign higher scores to responses that attached importance to growth and expansion of industries. On the other hand, love for tradition, contentment, and an easy way of life were given lower scores.

While scoring the responses concerning aspirations, the highest scores were given to those responses that emphasized doing things in a big way, for example, aspiring to own a big factory, be a great scientist, a big industrialist, a millionaire, etc. At the lower end were the responses that emphasized the routine way of doing things, for example, taking up a job, setting up a small stall or a shop, etc. Responses showing desire to close up business were scored still lower.

Interests. Here, persons showing no interest in anything other than their work were given a score of one. When the interests were confined to family affairs or circle of close friends, responses received a numerical score of two. The next higher category of responses involved activities without much intimate social contact (gardening, music, etc.). The highest scores were given to activities demanding more and more social contact (club, social and political activities, etc.). Thus, a person getting the lowest score has no time for anything other than his work, whereas a person with the highest score takes much active interest in social and political activities.

Administration of Tests

The entrepreneurs were interviewed by two trained psychologists in the months of May–September 1964. They had no information about the exponential growth rate of their *S*s. Their work was distributed in such a manner that each investigator had about 50% of the *S*s from each group. In general, two sessions were held for each *S*. For uniformity of procedure, the interview schedule and attitude scales were given in the first session, whereas the second session was devoted to sentence completion tests. The time required for the two sessions averaged about 2 hr.

In both the sessions, when preferred by *S*s, the responses were recorded at the appropriate places by the investigators. The responses were then converted into numerical values as explained earlier.

RESULTS AND DISCUSSION

The results have been discussed in terms of intergroup comparisons between metal and hosiery groups and intrasubgroup comparisons after dichotomizing the metal and hosiery groups at the median of their respective exponential growth rates. Both for the metal and hosiery groups similar intrasubgroup comparisons were made for those *S*s who had been administered the TAT. In the third phase the two groups were combined and then divided into fast- and slow-progress groups. The fast group was comprised of 92 *S*s, 46 each from the metal and hosiery groups with the exponential growth rate higher than their respective median. The other 92 *S*s comprised the slow-progressing group. The intercorrelations for the two groups were computed and factorized separately by the principal axis method.[4] In each case, four factors were extracted. The ex-

---

[4] Intercorrelation matrices for the fast and slow progressing groups have been deposited with the National Auxiliary Publications Service. Order Document No. 01474 from the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Science, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche, and make checks payable to Research and Microfilm Publications, Inc.

TABLE 2

MEANS, STANDARD DEVIATIONS, AND T RATIOS FOR INTERGROUP COMPARISONS

Test variable	Metal group		Hosiery group		t ratio
	$\bar{X}$	*SD*	$\bar{X}$	*SD*	
Job satisfaction	13.74	3.39	9.22	3.36	8.41**
Job prestige	18.55	2.70	18.69	3.54	.30
Value patterns	23.85	3.40	18.54	3.57	13.61**
Aspirations	33.85	4.68	28.55	6.50	5.29**
Interests	1.90	1.01	2.36	.92	4.18**
Achievement motivation	77.22	5.65	62.30	7.25	15.54**
Investment tendency	9.04	3.46	7.75	3.36	2.48*
Hoarding tendency	3.00	2.40	4.50	2.10	4.55**
Conflicting attitude	4.00	1.71	2.90	2.28	3.79**
Optimistic tendency	2.03	1.52	3.44	2.06	5.42**
Social work	3.10	1.50	3.50	1.45	1.82
Personal comfort	3.00	1.00	2.80	1.20	1.25
Attitude toward labor unions	4.59	1.57	5.60	1.40	4.81**

* $p = .05$, $t = 1.99$.
** $p = .01$, $t = 2.63$.

traction was done to the criterion of the Scree Test (Cattell, 1966). In the application of the Scree Test, before proceeding to the judgment on communalities and factor number, unities were inserted in the R matrix diagonal and latent roots were plotted down to eight factors separately for the two groups. The scree appeared at the fourth factor in each case. Communalities obtained for the four factors in the above analyses were used as guessed communalities in the extraction of factors. The factors were then rotated using the Varimax criterion for simple structure. The Varimax factors are reported in Table 5.

Inter- And Intracomparisons

Tables 2 and 3 report the means, standard deviations, and t ratios for the metal and the

TABLE 3

MEANS, STANDARD DEVIATIONS, AND T RATIOS FOR INTRAGROUP COMPARISONS

Test variable	Metal group					Hosiery group				
	Above *Mdn.*		Below *Mdn.*		t ratio[a]	Above *Mdn.*		Below *Mdn.*		t ratio[a]
	$\bar{X}$	*SD*	$\bar{X}$	*SD*		$\bar{X}$	*SD*	$\bar{X}$	*SD*	
Job satisfaction	15.30	3.20	12.18	2.80	4.89	10.00	3.10	7.44	2.50	5.96
Job prestige	19.45	2.70	17.65	3.15	2.90	18.50	3.25	18.88	3.79	.50
Value patterns	25.55	3.10	22.15	2.78	5.48	20.30	3.40	16.78	2.78	5.38
Aspirations	35.94	4.28	31.74	4.08	4.77	31.82	5.95	25.28	5.18	5.54
Interests	1.80	1.00	2.00	1.01	.01	2.20	1.00	2.52	.86	1.78
Achievement motivation	80.30	5.15	74.14	4.25	5.60	67.00	5.80	57.60	5.22	8.03
Investment tendency	8.88	3.34	9.20	3.56	.40	9.85	3.10	5.65	2.04	7.63
Hoarding tendency	2.60	2.30	3.40	2.43	1.60	4.30	1.90	4.70	2.26	.91
Conflicting attitude	4.35	1.50	3.65	1.83	2.00	3.00	2.22	2.80	2.34	.42
Optimistic tendency	2.00	1.48	2.06	1.56	.18	3.80	1.92	3.08	2.13	1.67
Social work	3.00	1.50	3.20	1.49	.64	4.10	1.30	2.90	1.34	4.86
Personal comfort	2.90	.94	3.10	1.05	1.00	2.40	1.00	3.20	1.25	3.33
Attitude toward labor unions	4.28	1.20	4.90	1.25	2.38	5.20	1.28	6.00	1.40	2.86

[a] At $p = .05$, $t = 2.01$; at $p = .01$, $t = 2.68$.

TABLE 4

MEANS AND MANN-WHITNEY *U* VALUES FOR THE ABOVE- AND BELOW-MEDIAN SUBGROUPS

Test variable	Metal group			Hosiery group		
	$\bar{X}$ above-*Mdn.* group	$\bar{X}$ below-*Mdn.* group	*U* value	$\bar{X}$ above-*Mdn.* group	$\bar{X}$ below-*Mdn.* group	*U* value
Job satisfaction	12.65	10.80	83*	12.00	5.20	58**
Job prestige	23.24	14.66	60**	21.38	15.62	53***
Value patterns	26.33	15.67	45***	22.50	14.43	53***
Aspirations	38.60	23.40	38	35.18	21.64	42***
Interest	1.90	2.00	90*	2.50	2.35	80*
Achievement motivation	85.13	54.13	28***	72.30	52.27	36
Investment tendency	13.00	6.20	47***	11.84	4.53	42***
Hoarding tendency	3.00	2.28	68*	5.00	4.30	63**
Conflict attitude	4.00	3.26	75*	4.22	2.30	70*
Optimistic tendency	2.35	3.00	70*	4.20	2.88	66*
Social work	5.15	2.00	46***	3.66	3.52	69*
Personal comfort	4.25	1.63	54***	3.38	2.83	74*
Attitude toward labor unions	3.00	5.62	58**	6.13	5.21	53***
Thematic Apperception Test	15.00	8.20	29	6.85	3.15	

*p = .05 (two-tailed test).
**p = .02 (two-tailed test).
***p = .002 (two-tailed test).

hosiery groups and also for the above-median and below-median subgroups in the metal and hosiery groups. In case of the above-median and below-median intrasubgroup comparisons (see Table 4) for *S*s who were administered the TAT, the significance of the differences was checked by the Mann-Whitney *U* test of two independent samples (Siegel, 1956).

A review of Tables 2 and 3 indicates that the metal group and the above-median subgroups differed significantly from their counterparts on job satisfaction, value patterns, aspirations, achievement motivation, and attitudes toward labor unions. The picture that emerges from the intracomparison of the smaller subgroups (see Table 4) is more or less comparable with the one obtained from their respective counterparts in the bigger samples. The minor differences can easily be attributed to sampling error. The differences, in all the cases, are in favor of the metal group and above-median subgroups. Keeping the scoring procedures in view, it implies that the fast-progressing group or subgroups tend to be more satisfied with their jobs, attach greater value to the importance of their growth and expansion of industries, want to do things in a big way, score higher on achievement motivation, and are more favorably inclined toward labor unions.

Further examination of the *t* ratios suggests that the metal and hosiery groups differ significantly on all the tested variables except those of job prestige, money spent on social work, and money spent on personal comforts. Such differences are inconsistent and much less marked in the subgroups. This is particularly true for Inv T, Hoa T, Conft A, and Opt T. This can be explained by assuming that these variables do not contribute much to entrepreneural success. The differences on these variables in the metal and hosiery groups are special characteristics of the two groups.

Factorial Study

Table 5 reports rotated factors for the fast- and slow-progressing groups, respectively. The common variance accounted for in the two analyses are 57.5% and 51.2% for the fast- and slow-progressing groups, respectively.

The major components of Factor 1, both in the fast- and slow-progressing groups are AM, Inv T, aspirations, value patterns, and Att L. It would appear from the description of thes

TABLE 5

VARIMAX-ROTATED FACTOR LOADINGS

Test variable	Fast-progressing group					Slow-progressing group				
	1	2	3	4	h^2	1	2	3	4	h^2
Job satisfaction	.22	.56	−.02	.19	.40	−.02	.63	.23	.01	.45
Job prestige	.46	.65	−.10	−.09	.65	−.34	.52	−.31	−.15	.51
Value patterns	.59	.12	.15	−.10	.40	−.60	.31	−.10	.09	.47
Aspirations	.76	.32	−.06	.11	.70	−.71	.20	−.01	−.02	.54
Achievement motivation	.81	.12	.02	.10	.68	−.72	−.07	.21	−.09	.57
Investment tendency	.69	.13	−.08	.01	.50	−.46	−.08	.07	−.09	.23
Hoarding tendency	.03	.10	.16	.92	.88	−.44	.24	−.04	.04	.25
Conflict attitude	−.02	−.01	.08	.01	.01	.05	−.01	−.03	.66	.44
Optimistic tendency	−.07	.09	.76	.14	.61	.07	−.01	.57	.66	.77
Social work	.12	.61	.56	−.24	.76	.08	.21	.79	−.08	.68
Personal comfort	.10	.67	.41	.13	.64	−.17	.47	.58	.32	.69
Attitude toward labor unions	.39	.25	.68	.02	.68	.23	.17	.67	.13	.55
% Contribution	20.84	14.88	13.31	8.48	57.51	16.66	9.72	16.12	8.72	51.22

tests that they measure tendency to strive, persist, cooperate, and go about things in a big way in business and industry. In view of these characteristics, Factor 1 is named AM. The structure of this factor in the slow-progressing group differs slightly as it has an additional significant loading on the variable of Hoa T. The contribution of Factor 1 in the fast- and slow-progressing group, is 21% and 17%, respectively.

Factor 2 has highest loading on the variables of job satisfaction, job prestige, and P Comf in both the groups. Further, the loadings on the variables of value patterns, aspirations, Soc W, and Att L indicate a similar trend in the two groups, but their values are rather low and nonsignificant in certain cases. Thus, the structure of this factor, too, is somewhat comparable in the two groups. Keeping in view the measurements made by the tests showing high loadings on this factor, it can be named the factor of self-image. Its contribution to the total variance in the fast- and slow-progressing groups is 15% and 10%, respectively.

The variables showing significant loadings on Factor 3 in the two groups are Opt T, Soc W, P Comf, and Att L. It appears difficult to find the common element that underlies the content of these variables. A tentative name that might be attached to this factor is living well. The contribution of this factor in the fast- and slow-progressing groups is 13% and 17%, respectively.

Factor 4 in the fast-progressing groups has significant loadings on the variables of Hoa T, and it practically accounts for the total variance attributed to this factor. In the slow-progressing group the significant loadings on this factor are on the variables of Conft T, Opt T, and P Comf. Apparently, the structure of this factor in the two groups is different. It is difficult to name this factor and may be regarded as quasi-specific.

The above discussion suggests that the structure of the three major motivational factors, in the two compared groups, is more or less comparable. Their contributions, however, differ: Factors 1 and 2 contribute more to the fast-progressing group, and Factor 3 contributes more to the slow-progressing group. Thus, it seems that the motivational makeup of the two groups, as explained by Factors 1, 2, and 3, differ quantitatively rather than qualitatively. On the fourth motivational factor, however, the difference is qualitative.

A review of the information provided by the factor structure and the one obtained through inter- and intracomparisons of groups shows that the two analyses supplement each other. It also reveals that the aspirations, AM, and Inv T are associated with a fast rate of industrial growth, whereas, Hoa T and Opt T seem to be related to the slow rate of industrial growth.

While the results obtained show the relationships of AM to other behaviors, they cannot prove that AM causes the differential rate of industrial growth characteristic of our study groups.

Since AM and rate of industrial development seem to be associated, it would be worthwhile to raise the level of AM of a group of people through counseling or special training and study its impact on economic growth in the underdeveloped countries. Pioneer work along these lines has already been initiated by McClelland (1964a, 1964b) and his team. Further studies in this direction need to be undertaken.

REFERENCES

Adams, J. F. Test item difficulty and reliability of item analysis methods. *Journal of Psychology*, 1960, 49, 255–362.

Bradburn, N. M. *The managerial role in Turkey: A psychological study*, Unpublished doctoral dissertation, Harvard University, 1960.

Bradburn, N. M., & Berlew, D. E. Need for achievement and English industrial growth. *Economic Development and Cultural Change*, 1961, 10, 8–20.

Cattell, R. B. The Scree Test for the number of factors. *Multivariate Behavioral Research*, 1966, 1, 245–276.

Cortes, J. B. The achievement motive in the Spanish economy between 13th and 18th centuries. *Economic Development and Cultural Change*, 1961, 9, 144–163.

Hagen, E. M. Psychology's role in economic development. *Science*, 1961, **134**, 1608–1609.

McClelland, D. C. *The achieving society*. New York: Appleton-Century-Crofts, 1961.

McClelland, D. C. The achievement motive in economic growth. In B. F. Hoselitz & W. E. Moore (Eds.), *Industrialization and society*, Paris: UNESCO, 1963.

McClelland, D. C. *Changing values of progress*. Unpublished manuscript, Harvard University, 1964. (a)

McClelland, D. C. *Toward a theory of motive acquisition*. Unpublished manuscript, Harvard University, 1964. (b)

Rangnekar, B. S. *Small industries and social change*. New Delhi: UNESCO, 1966.

Rogers, E. M., & Neill, R. E. *Achievement motivation among Columbian peasants*. (Diffusion of innovation research report) East Lansing, Michigan: Michigan State University, 1966.

Rohde, A. R. Explorations in personality by the sentence completion method. *Journal of Applied Psychology*, 1946, **30**, 169–181.

Rostow, W. W. *The stages of economic growth*. New York: Cambridge University Press, 1961.

Rotter, J. B., & Willerman, B. The incomplete sentences test as a method of studying personality. *Journal of Consulting Psychology*, 1947, **11**, 43–48.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Thurstone, L. L., & Chave, E. J. *The measurement of attitude, a psychological method*. Chicago: University of Chicago Press, 1929.

Weber, M. *The protestant ethic and the spirit of capitalism*. New York: Charles Scribner's Sons, 1930.

(Received May, 5 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 324–330

EFFECTS OF MODE OF ORGANIZATION AND FEEDBACK LEVEL ON CREATIVE TASK GROUPS[1]

THOMAS H. STONE[2]

Department of Business Administration, University of Iowa

A laboratory experiment tested the effects of mode of organization (specialization versus project teams) and feedback level (individual versus group) on the performance and satisfaction of peer-evaluated, creative task groups. Twenty-four groups of three *S*s each worked on a word-construction task. Performance on each trial of the task was measured by three indexes: score on the final word constructed (word score), time to achieve a task solution, and rated creativity of the word constructed. An index of task satisfaction was obtained from questionnaire items. Results showed that creative task groups organized into project teams had shorter times and expressed higher satisfaction with the task than did groups organized by specialization. Groups receiving individual-level feedback had higher mean word scores than did groups receiving group-level feedback.

A common source of many of today's inventions and product improvements is the creative task group, or research and development (R&D) group. Relatively little research has been done on such creative task groups, and what has been done has tended to be in ongoing organizations using questionnaires and interview techniques. Pelz and Andrews' (1966) study of 1,300 scientists and engineers in 11 R&D laboratories represents one of the few comprehensive studies of this type. Almost no experimentation has been done with R&D groups, perhaps due to the specialized, complex, and valuable nature of R&D functions. Experimentation with ongoing organizations is always subject to difficulties, among them, subject mortality, uncontrolled (and often uncontrollable) factors, limited generality of the findings, and high cost to both researcher and the organization under study.

On the other hand, excellent arguments have been made favoring the laboratory as a site for testing organization hypotheses (see Weick, 1965; Festinger, 1953). Carzo (1963), Carzo and Yanouzas (1969), and Becker and Baloff (1969) have conducted laboratory experiments in which the effects of organization structure were examined. Another laboratory experiment (Zander & Wolfe, 1964) used members of existing managerial committees to examine the effects of individual and group-level feedback on performance and morale. In the research reported here, the laboratory was used as a testing ground for examining hypotheses regarding the effects of mode of organization (specialization versus project teams) and feedback level (individual versus group) on the performance and satisfaction of creative task groups.

Since creative task groups are usually organized specifically for the purpose of generating new or unique ideas or products, it is important that they be organized and managed in ways that are most conducive to high group productivity. The problem of how research laboratories should be organized is one of nine dilemmas in industrial research posed by Shepard (1956). Shepard suggests that there are two basic modes of organization, specialization and "organization by task." In the specialization mode of organization, each person in the group approaches the task as a specialist or specializes in only one part of the task. The specialization mode is frequently characterized by a clear division of labor among members of the creative task group. Project teams typify the "organization by task" mode. Such groups are organized around a certain task, their raison d'être being to solve a particular problem or class of problems. Group members organized in this manner do not seek to dissect the group problem into their respective areas of specialty; rather, they attack the problem or task *as a group*.

---

[1] This research is based on the author's doctoral dissertation in the Psychology Department of the University of Minnesota.

[2] Requests for reprints should be sent to Thomas H. Stone, Department of Business Administration, University of Iowa, Iowa City, Iowa 52240.

Literature comparing the two modes of organization has tended to favor project teams over groups organized by specialization. Shepard (1956), for example, has stated that the specialization mode leads to a number of difficulties such as "resistance to crossing specialist boundaries and cooperating with other groups. . . and a lack of creativity and responsiveness to challenge [p. 302]." In his study of 11 R&D laboratories, Pelz (1967) found that high-performing scientists tended to be those with not one, but two or three areas of specialization. He also concluded that "specialization lends security, but diminishes challenge [p. 165]." On the other hand, it has been said that groups organized into project teams will often turn out to be "enthusiastic, creative, and dramatically successful [Shepard, 1956, p. 302]" and also more innovative than those groups emphasizing "narrow, nonduplicating, nonoverlapping definitions of duties and responsibilities [Thompson, 1965, p. 13]."

Let us now turn our attention to the second variable under investigation in this study—feedback level. The problem of reward and recognition is another of Shepard's (1956) nine dilemmas in industrial research. To the extent that a reward system provides a person or a group with knowledge of the relative quality of a given performance, a reward system is a feedback system. Feedback may be administered on either an individual or a group level; that is, feedback may be directed toward each person in an organizational unit (individual-level feedback) or toward the organizational unit as a whole (group-level feedback).

Literature discussing individual versus group-level feedback has tended to favor the former. Group-level feedback does not furnish the individual with knowledge of his own personal contribution to a group product. The absence of job performance feedback to the individual is a frequent complaint of organization members according to Katz and Kahn (1966). Indeed, individual performance evaluations have been viewed as important to "the development and maintenance of one's self-concept [Scott, Dornbusch, Busehing, & Laing, 1967, p. 95]." Shepard (1956) has stated that "the reward system should be sensitive to and accurate in response to individual task achievement [p. 305]." By responding to individual task achievement, the reward system may serve to motivate the individual to further accomplishments.

Performance evaluations are necessary if feedback is to be administered to individuals or groups. In the case of creative task groups, two sources of evaluation are common. Evaluation of performance may be undertaken either by a single supervisor who is not a member of the group or by task group members themselves. The latter alternative, "peer evaluation," has become increasingly popular in recent years. As a method of performance appraisal, it has been used successfully in both military situations (Hollander, 1954; 1965) and industrial settings (Roadman, 1964; Weitz, 1958). Peer evaluation was employed by the 24 creative task groups in this study.

The following hypotheses were tested:

1. *Performance and satisfaction of groups organized into project teams will be higher than those organized by specialization.*

2. *Performance and satisfaction of groups given individual-level feedback will be higher than those receiving group-level feedback.*

METHOD

Design. In this experiment, mode of organization and feedback level were examined in a 2 × 2 design that resulted in four cells: project—individual; project—group; specialization—individual; and specialization—group. Each cell contained 6 groups, making a total of 24 experimental groups. Each group had three *S*s who worked together on a word construction task. The experiment was pretested and minor adjustments in the manipulations resulted.

Subjects. The *S*s were 72 students in elementary psychology courses at the University of Minnesota. Students who volunteered to take part in the 2-hr. experiment received bonus points for their participation.

Task. A word-construction task devised by *E* was used. The object of the task was to *make as creative and as high scoring a word as possible in the shortest amount of time*, using a number of letters and employing three rules. There were no correct or preplanned solutions to any of the five trials of the task.

For each trial, six letters were placed face up in alphabetical order on a table. In addition six letters were placed in a row, face down on the table. The backs of these unknown letters were numbered 1, 2, 3, 4, 5, and 6. Through the use of three rules, *S*s could either: (*a*) set aside permanently one of the face-up letters and replace it with a letter drawn (in numerical order) from the row of unknown letters (Rule No. 1); (*b*) add to the face-up letters a letter drawn (in numerical order) from the unknown letters (Rule No. 2); or (*c*) chang one of the face-up letters into any letter of their choice (Rule No. 3). Use of a rule entailed a cost that varied according to utility of the rule (i.e., Rule No. 3 cost more to use than did Rule No. 1). An 8-min. time limit was set on each trial, with warnings given at 6 and 7 min.

Each trial was scored by counting the number of letters in the final word constructed by the group, multiplying this number by 20, and then subtracting the number of points entailed by use of rules. The resulting number for each trial was a group's final word score.

Manipulations. Groups were organized by specialization or into project teams by having *S*s in each mode approach the task in a different way. Regardless of mode, the same three rules were used to work the task. However, in groups organized by specialization, each *S* was given a different rule to work with (i.e., Person No. 1 received Rule No. 1, etc.) while *S*s organized into project teams were each given the same three rules to work with (i.e., each *S* received all three rules). Each *S* in the specialization mode was told that he was to be considered an expert on the use of his particular rule. These *S*s were also instructed: "In working on the task *you are to advocate the use of your rule and actively encourage the other group members to use it.* Remember, you will be rated on how well you do this." In the project team mode, however, *S*s given the same three rules to work with were instructed: "In working on the task *you are to cooperate with one another.* Remember, you will be rated on how well you do this."

To manipulate level of feedback, *S*s were told that they would be receiving either individual or group-level feedback in the form of performance evaluation scores. Individual performance evaluation scores informed each *S* of how his performance compared to each of the other two members in his group. Group performance evaluation scores let each *S* know how his group's performance compared to that of two comparison groups.[3] The *S*s in each feedback condition were told that "a high performance evaluation score reflects both high verbal ability (a part of intelligence) and creativity."

Since peer evaluation was employed, *S*s were also told that they would be evaluating either individual or group performance. In individual feedback conditions, each *S* rated the performance of each of the other two members in his group. In group feedback conditions, each *S* rated his group as a whole. The *S*s were told that performance evaluation scores were averages of the ratings they assigned to one another or to the group. Actually, *S*s received false feedback scores.

Procedure. After reading their instructions and rule sheets, *S*s were provided with an example of how the task was worked and scored. They were then allowed one practice trial, after which the series of five trials began. Immediately following each trial, *S*s were given brief rating sheets and asked to independently evaluate performance on the trial just completed. Rating sheets were collected by *E* as soon as the *S*s had finished filling them out. Feedback was delayed one trial, that is, performance evaluation scores for Trial 1 were received only after *S*s had finished evaluating their performance on Trial 2, etc. Feedback was lagged in this way both to enhance the credibility of false feedback and to simulate the delay of knowledge of results common to R&D groups. A one-page midexperiment questionnaire was filled out by *S*s after completion and evaluation of Trial 2 but before any feedback from Trial 1 was given. This questionnaire was designed to provide a measure of satisfaction with the task. It was administered before any feedback was given in an attempt to avoid possible contamination due to such feedback. After receiving feedback from Trial 5, *S*s filled out a short postexperiment questionnaire and were debriefed.

Measures of Performance

Three measures of task performance were used: word score, creativity, and time.

Word score. High word scores on the task indicated good performance. Word score was calculated for each of the five trials by multiplying the number of letters in the constructed word by 20 and then subtracting the number of points accrued through use of rules. Mean word score, over five trials, for the 24 groups was 90.79, with a standard deviation of 6.71.

Creativity. High creativity scores indicated good performance on the task. This performance measure was obtained from 7-point Likert-scale ratings of creativity. A panel of three judges was told briefly about the word-construction task. Each judge was then asked to rate as to creativity the many different words constructed by the 24 groups over all trials. Average interrater reliability for creativity ratings was .71. Creativity scores were assigned to each word by averaging the three creativity ratings given it. Mean creativity, over five trials, for the 24 groups was 3.59, with a standard deviation of .78. Creativity correlated .42 with word score.

Time. Good performance with respect to time was indicated by fewer seconds taken to complete each trial of the task. The time limit set for each trial was 8 min., or 480 sec. Mean time, over five trials, for the 24 groups was 313.96 sec. with a standard deviation of 72.93 sec. Time and word score correlated .17; time and creativity correlated .12.

Task satisfaction measure. Task satisfaction was an index of five questionnaire items, four of which were 5-point Likert-scale items. The items touched on how *S*s felt while working on the task, how much they liked it, and how much interest and fun it held for them. Three of the task satisfaction items used were from the midquestionnaire, which was administered after Trial 2 and before *S*s received any performance feedback. The other two items used were from the postquestionnaire. The Kuder-Richardson internal consistency reliability (see Nunnally, 1967, pages 195–196) for the items used in this index was .74. Mean task satisfaction for the 24 groups was 18.30, with a standard deviation of 1.66. Task satisfaction correlated − .16 with word score, − .08 with creativity, and − .56 with time.

---

[3] Two comparison groups were used in the group feedback condition so that the same false feedback scores could be used for both individual and group feedback conditions.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF THE FOUR DEPENDENT VARIABLES ON MODE OF ORGANIZATION AND FEEDBACK LEVEL

Dependent variable	Mode of organization				Feedback level			
	Project team		Specialization		Individual		Group	
	X	SD	X	SD	X	SD	X	SD
Time	285.50	57.91	342.42	71.03	322.67	58.67	305.25	70.39
Task satisfaction	19.13	1.38	17.47	1.24	18.44	1.25	18.17	1.37
Word score	92.50	5.63	89.08	5.72	93.17	6.90	88.42	4.09
Creativity	3.54	.75	3.63	.76	3.49	.77	3.68	.74

RESULTS

Separate analyses of variance were done on each of the four dependent measures in the study: word score, creativity, time, and task satisfaction. With the exception of creativity, a significant effect was found on each dependent measure. Means and standard deviations of the four dependent variables on mode of organization and feedback level are given in Table 1.[4]

Regarding mode of organization, it was hypothesized that performance and satisfaction of groups organized into project teams would be higher than those organized by specialization. Significant effects of mode on two of the dependent measures, time and task satisfaction, supported this hypothesis.

The effect of mode on time was significant at the .06 level. As seen in Table 1, groups organized into project teams had a mean time of 285.50 sec., while groups organized by specialization on the average required more time (342.42 sec.). This difference accounted for 11% of the total variance for time.

The effect of mode on task satisfaction was significant at less than the .01 level. This effect, which accounted for 18% of the variance, supported the hypothesis of higher satisfaction in groups organized into project teams. Mean task satisfaction for groups organized into project teams was 19.13 as compared to 17.47 for groups organized by specialization (see Table 1).

In light of the above findings, it is interesting to note the significant correlation ($r = -.56$) between time and task satisfaction (see Table 2). This correlation shows that low mean times are associated with high satisfaction scores.

There were no significant effects of mode of organization on either word score or creativity.

Regarding feedback level, it was hypothesized that performance and satisfaction of groups given individual-level feedback would be higher than those receiving group-level feedback. Although there was no significant effect of feedback level on satisfaction, findings for the variable word score lent support to the hypothesis regarding performance.

The effect of feedback level on word score was significant at the .07 level. Groups given individual-level feedback had higher mean word scores than did groups receiving group-level feedback (93.17 as compared to 88.42,

TABLE 2

CORRELATIONS BETWEEN THE DEPENDENT VARIABLES

Dependent variable	1	2	3
1. Word score	—		
2. Time	.17	—	
3. Creativity	.42*	.12	—
4. Task satisfaction	−.16	−.56*	−.08

Note.—N = 24 groups.
* $p < .01$.

---

[4] ANOVA tables for the variables time, word score, and task satisfaction have been deposited with the National Auxiliary Publications Service. Order Document No. 01475 from the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

see Table 1). This difference accounted for 9% of the variance.

Table 2 shows that word score is highly correlated with creativity, a measure for which no significant results emerged. Word score was not significantly correlated with either time or task satisfaction.

There were no significant effects of feedback level on time, creativity, or task satisfaction.

DISCUSSION

This study is only a preliminary step toward an understanding of how organizational structure and feedback level may affect the complex workings of "real life" creative task groups. Further field and laboratory research are certainly necessary before R&D group managers modify their practices. Admittedly, elements in the laboratory situation are not always analogous to conditions existing in the larger organization, and it is difficult to estimate the degree to which the manipulations in this experiment accurately reflect crucial properties in ongoing R&D groups. In the first place, the study made no attempt to build in effects of extended contact between group members, such as norms, standards, and status differentiations. The specialization manipulation of giving each *S* a different rule to advocate may not have been sufficiently powerful to capture the primary characteristics of "real" specialists. And, undoubtedly, the personal consequences of performance feedback—whether individual or group—were not as great for the experimental *S*s as they would have been for actual R&D group members whose job security may be affected. Furthermore, the experimental task may have somewhat limited generalizability due to its essentially verbal nature. With these limitations in mind, let us proceed to a discussion of the findings.

Performance of R&D groups, or creative task groups, if often evaluated according to multiple criteria, among them product quality, rapidity of problem solution, and creativity or product uniqueness. The three measures of performance used in this study—word score, time, and creativity—were designed to be analogous to actual R&D performance criteria. The importance of using multiple criteria of performance in research studies on R&D groups is illustrated by the results of this study. Different performance criteria were not affected uniformly by the two factors under study: mode of organization affected time, but not word score; feedback level affected word score, but not time. This finding may have implications for the organization and management of R&D groups. When rapidity of problem solution is an important criterion, attention might be given to the mode of organization factor, feedback level being of little consequence. However, if product quality (here indicated by word score) is of the utmost importance, attention might be paid to the feedback level factor, mode of organization not being a major concern.

Specifically, it was found that groups organized into project teams tended to find solutions faster than did groups organized along specialist lines. The *S*s in the specialization mode, each approaching the task as an "expert" and advocating the use of one particular rule (or approach), required more time. Since a single task solution was required from each group for each trial, *S*s in specialization groups may have needed more time to resolve differences in viewpoint and to integrate dissimilar approaches into a common task solution. Groups organized into project teams did not have these problems, since each group member was given all three rules with which to work in a more cooperative venture.

Groups organized into project teams also tended to express higher satisfaction with the task than did specialization groups. Perhaps *S*s organized into project teams enjoyed the cooperative nature of their approach to the task. Or maybe they liked the freedom they had with which to work, not being restricted to advocating one particular rule or approaching the problem from a certain viewpoint. The situation may have been more stressful for groups organized by specialization. This specialization involved groups whose members had vested interests which perhaps caused them to work against one another to ensure that their particular rule was used.

Another possible explanation for the higher task satisfaction of project teams is that these groups may have been more satisfied with the task because of their success at it in terms of the performance criterion time. That people tend to gain satisfaction from success has been pointed out by Locke and Breer (1965). This explanation would suggest that one's satisfaction with the task itself may not be completely separate from how well one does, or thinks he does, on the task. Task satisfaction and time were correlated − .56: satisfaction is high when time is low and vice versa. Due to the fact that performance feedback to *S*s was lagged one trial, *S*s perhaps formed their own, more immediate ideas of performance quality after completing each trial. Of the three performance criteria, time would have been the most obvious for *S*s; the 8-min. time limit posed an absolute standard to work against. (Creativity is a fairly ambiguous concept, and there was no absolute word score for *S*s to strive for.) If, indeed, *S*s did form their own ideas of performance quality, they may have felt more successful (and hence more satisfied) the less time they took.

Feedback level affected only word score. Individual-level feedback was found to produce higher mean word scores than did group-level feedback. If one accepts word score as an indicator of quality, this finding suggests that higher quality solutions or products may result when R&D group members' contributions are individually recognized. This supports the notions of Scott et al. (1967) and Shepard (1956) that individual-level feedback is preferable to group-level feedback.

Feedback level did not affect satisfaction with the task. Perhaps the reason for this was that three of the five items in the task satisfaction index were obtained before *S*s had received any feedback whatsoever. Therefore, for the most part, they had had no real experience with either individual- or group-level feedback.

The failure of the performance criterion creativity to produce any significant effects may be due to the nature of the creativity measure. The measure itself may not be valid, since ratings of constructed words were made by persons who had not performed the task. An alternative explanation is that the measure is valid, but that neither mode of organization nor feedback level has a differential effect on creativity.

Further research might focus on the combined effects of mode of organization and level of feedback in the hope that a combination could be found that might maximize all three performance criteria—high quality, high creativity, and low time. Other measures of creativity should be used in such studies. Research should also be extended to the field to examine the effects of these variables on the performance and satisfaction of ongoing R&D groups. Finally, since peer evaluation is still relatively rare in R&D groups, further research should include evaluation by a superior or group supervisor.

Summary. In brief, groups organized into project teams tended to complete the trials faster than did groups organized by specialization. Groups organized into project teams also tended to express higher levels of task satisfaction than did those organized by specialization. Regarding feedback level, groups given individual-level feedback tended to construct higher scoring words than did groups receiving group-level feedback. No significant effects were found on the performance variable creativity.

REFERENCES

BECKER, S. W., & BALOFF, N. Organization structure and complex problem solving. *Administrative Science Quarterly*, 1969, **14**, 260–271.

CARZO, R., JR. Some effects of organization structure on group effectiveness. *Administrative Science Quarterly*, 1963, **7**, 393–424.

CARZO, R., JR., & YANOUZAS, J. N. Effects of flat and tall organization structure. *Administrative Science Quarterly*, 1969, **14**, 178–191.

FESTINGER, L. Laboratory experiments. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Holt, Rinehart & Winston, 1953.

HOLLANDER, E. P. Peer nominations on leadership as a predictor of the pass-fail criterion in naval air training. *Journal of Applied Psychology*, 1954, **38**, 150–153.

HOLLANDER, E. P. Validity of peer nominations in predicting a distant performance criterion. *Journal of Applied Psychology*, 1965, **49**, 434–438.

KATZ, D., & KAHN, R. *The social psychology of organizations*. New York: Wiley, 1966.

LOCKE, E. A., & BREER, P. C. *Task experience as a source of attitudes*. Homewood, Ill.: Dorsey Press, 1965.

NUNNALLY, J. C. *Psychometric theory*. New York: McGraw-Hill, 1967.

PELZ, D. C. Creative tensions in the research and development climate. *Science*, 1967, **157**, 160–165.

PELZ, D. C., & ANDREWS, F. M. *Scientists in organizations*. New York: Wiley, 1966.

ROADMAN, H. E. An industrial use of peer ratings. *Journal of Applied Psychology*, 1964, **48**, 211–214.

SCOTT, W. R., DORNBUSCH, S. M., BUSEHING, B. C., & LAING, J. D. Organizational evaluation and authority. *Administrative Science Quarterly*, 1967, **12**, 93–117.

SHEPARD, H. A. Nine dilemmas in industrial research. *Administrative Science Quarterly*, 1956, **1**, 295–309.

THOMPSON, V. A. Bureaucracy and innovation. *Administrative Science Quarterly*, 1965, **10**, 1–20.

WEICK, K. E. Laboratory experimentation with organizations. In J. G. March (Ed.), *Handbook of Organizations*. Chicago: Rand McNally, 1965.

WEITZ, J. Selecting supervisors with peer ratings. *Personnel Psychology*, 1958, **11**, 25–35.

ZANDER, A., & WOLFE, D. Administrative rewards and coordination among committee members. *Administrative Science Quarterly*, 1964, **9**, 50–69.

(Received April 22, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 331–338

DEVELOPMENT OF THE SURVEY OF WORK VALUES

STEPHEN WOLLACK,[1] JAMES G. GOODALE, JAN P. WIJTING,

AND

PATRICIA CAIN SMITH[2]

Bowling Green State University

A series of scales measuring attitudes toward work was developed. The Survey of Work Values (SWV) was based on a number of dimensions of Protestant Ethic, specifically those aspects that deal with the meaning that an individual attaches to his role at work. The content validity of the items was established by the reallocation method of scaling. Items were analyzed to determine the best method of scoring and to select items for the subscales according to the internal consistencies and independence of the subscales. Unweighted, multipoint scoring was selected for 54 items. SWV scores discriminated among occupational groups and correlated with background characteristics of employed and disadvantaged persons.

The purpose of this research was to construct a set of scales measuring several areas of work values. The Survey of Work Values (SWV) is intended to be an index of a person's attitudes toward work in general, rather than his feelings about a specific job. The concept of work values, referring to general attitudes regarding the meaning that an individual attaches to his work role, therefore, differs from that of job satisfaction (an attitude toward one's own job).

A few scales have been constructed for the purpose of measuring work values (Stefflre, 1959; Super, 1957, 1962) and occupational values (Kilpatrick, Cummings, & Jennings, 1964; Rosenberg, 1957). Although these measures have been carefully developed, they seem to be extremely global. The SWV differs from previous scales[3] in that it is directed toward separate areas of values and is limited to the construct of secularized Protestant Ethic with which work values seem to be closely linked.

The principal aspects of Protestant Ethic as described by Weber (1958) are individualism, asceticism, and industriousness. The emphasis placed on a man's industriousness probably represents the most critical aspect of Protestant Ethic. The Ethic has been assessed typically by indirect methods that have been presumed to index this concept. Amount of industrialization, amount of education, economic level, and occupation have been among the more frequently used indexes.

Correlations between attitudes and supposedly logically related behaviors have usually been found to be low. Frequently, a variety of considerations intervene to inhibit the behavioral manifestations of an attitude. Economic and social factors may greatly limit the alternative behaviors available to an individual regardless of his attitudes. It would, therefore, be naïve to expect indirect measures to index accurately an individual's work values. Attitude scales should provide more direct measures of these concepts.

The categories of the SWV were selected on the basis of two considerations. The first was the apparent relevance of each category to the secularized interpretation of Protestant Ethic, specifically those aspects dealing with work. Although this selection was judgmental, it did narrow considerably the number of dimensions by eliminating the primarily religious aspects of the Ethic. A second consideration was that the categories should be relevant to the literature concerning the Protestant Ethic.

Despite an effort to choose areas of work values that provide a broad interpretation of Protestant Ethic, the categories of the SWV are by no means exhaustive. Unavoidably, some question may exist regarding any particular aspect of Protestant Ethic that was not included among the subscales. Practicality, however, dictates a relatively brief scale.

---

[1] Now at the Procter & Gamble Company.

[2] Requests for reprints, the items, their scale values, and permission to use the scales for research purposes may be obtained by writing Patricia Cain Smith, Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43402.

[3] The authors are familiar with only one previous effort to develop scales of work values based on Protestant Ethic (Blood, 1969). His scales consist of two dimensions, each containing four items measuring "pro-Protestant Ethic" and "non-Protestant Ethic."

Probably the most widely accepted notion of Protestant Ethic deals with the intrinsic aspect of work; that is, work as its own reward. Work is to be valued because it represents the best use of a man's time, not merely because it is instrumental to the attainment of external rewards. The employee high in Protestant Ethic is presumed to prefer working to being idle, to be involved in his work, and to derive considerable satisfaction from doing his job well. The authors selected three dimensions of Protestant Ethic that cover the intrinsic aspects of work:

Pride in Work: the satisfaction and enjoyment a man feels from doing his job well.

Job Involvement: the degree to which a worker takes an active interest in co-workers and company functions and desires to contribute to job-related decisions.

Activity Preference: a preference by the worker to keep himself active and busy on his job.

Although the traditional Ethic stressed the intrinsic rewards of work, considerable value was placed on extrinsic rewards as well. The following subscales reflect the extrinsic nature of the Ethic:

Attitude toward Earnings: the value an individual places in making money on the job.

Social Status of Job: the effect the job alone has on a person's standing among his friends, relatives, and co-workers, in his own eyes and/or in the eyes of others.

Two dimensions of the Ethic have been included that do not conform well to the intrinsic-extrinsic dichotomy and were, therefore, regarded to be of a mixed character.

Upward Striving: the desire to seek continually a higher level job and a better standard of living.

Responsibility to Work: the recognition that a man is obligated to work and that he must depend on himself rather than others for support.

METHOD

General definitions, as well as more specific items epresenting each subscale, were written on the basis of review of the literature and suggestions from faculty nd graduate students, all having had industrial experi- nce. The representativeness of these categories was then checked with an industrial sample. Generality was investigated in two undergraduate samples. Scale values were obtained on another industrial sample and an undergraduate sample. Endorsements of the items were then obtained on still another industrial sample.

Scaling by Method of Reallocation

First, the a priori attitude statements and the list of defined categories were presented to a group of 58 employees of a glass-manufacturing company in the Midwest. The workers ranged from the unskilled to the managerial level. In accordance with the method of reallocation described by Smith and Kendall (1963), the judges were instructed to assign each of the 91 statements into the defined category that they believed best corresponded with the meaning of the statement. In addition to the seven defined categories, a category designated as other was included. The judges were instructed to allocate to this category any statement that they believed to correspond poorly with each of the defined categories. The purpose of the reallocation procedure was to determine: (*a*) whether discriminably different categories exist, (*b*) whether the items intended to measure a particular concept are judged to be relevant to that concept, and (*c*) whether the investigator's conception of the categories being measured corresponds closely with the judgment of persons for whom the scale is intended.

The criteria for the retention of an item were at least 70% allocation to a single category and no more than 20% allocation to any second category. An item was rejected if it failed to meet either of these criteria. A scale was eliminated if it was not conceptually clear, as was indicated by allocation of items to other than the intended scale. Retained items were submitted to two further reallocations by samples of 56 and 57 undergraduates. Any item that failed to meet the criteria on any one reallocation was deleted. The use of a nonindustrial population was intended to establish the stability of these judgments over distinctly different groups.

Then scale values indicating level of each statement along the relevant continuum were obtained from a sample of 46 workers from the same company representing the same range of occupational levels. The judges were given six lists of statements, each headed by the category name and definition. (The Responsibility subscale had been eliminated; see Results section.) The rating scale consisted of five letters with a different pair of behavioral descriptions as end anchors for each of the six scales. The points of the scale were described as hypothetical workers. The judges were instructed to indicate which of the five hypothetical workers was most likely to have made each of the statements. Scale values were computed on the basis of a 5-point rating scale.

To substantiate the scale values, which had been obtained from a relatively small group of judges, the statements were rescaled by a group of 45 undergraduates, and the two sets of values were correlated. Since the two groups of judges differed markedly in their backgrounds, the correlation between the two sets of scale values reflects the stability of the relative positions of these ratings across different populations rather than their reliability across two samples drawn from the same population.

The 67 attitude statements were presented in mixed order to a group of 495 employees from three plants of the glass-manufacturing company mentioned previously. Employees were instructed to respond to each item on a 6-point scale of agreement ranging from strongly agree to strongly disagree. The sample included seven occupational groups: management ($N = 35$), supervision ($N = 79$), professional ($N = 82$), technicians ($N = 37$), clerical ($N = 120$), skilled trades ($N = 47$), and semiskilled and unskilled ($N = 85$). Also included in the sample were 10 workers for whom job titles were missing.

Scoring of the Survey of Work Values

Strength of agreement was computed from a 6-point Likert scale, which was also scored dichotomously (agree–disagree). Agreement values were reflected in order to maximize the number of positive intercorrelations within each subscale. For unweighted multipoint and unweighted dichotomous scoring, agreement responses were simply summed for each subscale. Weighted multipoint scoring and weighted dichotomous scoring summed the products of the agreement value of each statement with its corresponding scale value.

Product-moment correlations were computed for each item against total score for the subscale, and all of the items were intercorrelated. Also, subscale totals were correlated both within and across the four methods of scoring. Because each correlation between item and subscale total was based on a small number of items, a correction for the possible spurious effect of autocorrelation was made (see Nunnally, 1967, p. 262). For each subscale, items were arranged in order of their correlation with total subscale score. Coefficient alpha (a measure of internal consistency) was computed for the items with the highest item-total correlations. Other items were added successively to this core of items in order of the magnitude of their correlations with total score. Items were retained only if their addition resulted in an increment to alpha. Preference was given to items spaced widely along the subscale continuum, but with scale values of low variability.

An item was retained, therefore, if it: (*a*) met reallocation standards, (*b*) had a high correlation with its subscale total score, (*c*) resulted in an increment to coefficient alpha, (*d*) increased the distribution of scale values among items of its subscale, and (*e*) had a scale value of low variability.

Dimensionality of the Survey of Work Values

The dimensions obtained by reallocation need not correspond with those obtained by factor analysis, especially since reallocation is based on discriminability among the scales, rather than orthogonality. Scoring by factors might be better than scoring by allocation subscales. Therefore, the factor structure of the SWV was examined. The items administered to the last industrial sample ($N = 495$) were analyzed by a principal components analysis with quartimin rotation (Goodale, 1969b).

RESULTS

Of the 91 items in the original pool, 45 were retained after allocation. They had been assigned to a particular category on an average of 84.1% of the judgments. The Responsibility to Work subscale was eliminated because its items did not meet allocation standards. The Activity Preference items also failed to meet the stated criteria because the items were consistently allocated to the Responsibility to Work scale as well as to the intended scale. The definition for Activity Preference was, therefore, revised and additional items were written.

The 45 items retained from the first allocation, plus the 13 statements for Activity Preference that had failed to allocate, as well as 17 new statements, were subjected to two subsequent allocations. Two samples, consisting of 56 and 57 undergraduate students, were instructed to allocate the statements into the six defined categories and an "other" category. Of the 75 statements, a total of 67 were retained with an average value of 85% allocation to a single category. A total of 42 statements had survived two reallocations, while 25 statements had been reallocated once. Of the latter 25 statements, 12 were from the Activity Preference scale, and 13 were from the new items that were added to the item pool.

Scale Values

The rating of item scale values by the industrial sample produced a wide range of mean values within each of the six subscales, with relatively low standard deviations. The correlation between the two sets of scale values obtained from the industrial and undergraduate samples was .94 ($N = 67$). In addition, no systematic displacement of these ratings had occurred.

Total subscale scores based on the 6-point ratings of agreement were highly correlated ($r = .98$) with totals based on the summed products of the agreement weights and scale values. Dichotomous scoring of the unweighted values was also strongly related ($r = .96$) to the totals derived from the multiplicative approach. The unweighted multipoint scoring

TABLE 1

ITEM ENDORSEMENT MEANS AND STANDARD DEVIATIONS

Subscale	Range of $\bar{X}$s	Range of *SD*s
Status	2.32–3.88	.95–1.64
Activity	3.77–5.60	.62–1.28
Striving	3.47–5.54	.55–1.36
Earnings	2.01–4.41	.99–1.33
Pride	4.86–5.72	.55–1.04
Involvement	4.16–5.42	.63–1.34

Note.—$N = 495$.

method was retained because it increased item variance over that from dichotomous scoring.

Scale Characteristics of the Survey of Work Values

The item analysis resulted in six subscales each containing nine items. Table 1 summarizes the means and standard deviations for endorsement of the items of the SWV.

Median intrascale item intercorrelations and reliabilities (alpha) are shown in Table 2. To substantiate these values, the SWV was also administered to a sample of government employees ($N = 356$) and intrascale item intercorrelations and alpha coefficients were recomputed. The results of the latter analysis (also shown in Table 2) agree closely with the values obtained from the development sample.

The subscale reliabilities are considered to be satisfactory in view of the relatively small number of items within each subscale. Also, alpha tends to be small when item means are as extreme as those shown in Table 1.

TABLE 2

MEDIAN INTRASCALE ITEM INTERCORRELATIONS, COEFFICIENT ALPHA RELIABILITIES, AND TEST–RETEST RELIABILITIES

Subscale	Industrial workers[a]		Government workers[b]		Insurance employees[c]
	Mdn. r	Alpha	*Mdn. r*	Alpha	Test-retest
Status	.16	.63	.12	.55	.71
Activity	.16	.63	.15	.61	.71
Striving	.14	.59	.12	.55	.76
Earnings	.16	.63	.18	.66	.65
Pride	.16	.63	.15	.61	.69
Involvement	.11	.53	.16	.63	.68

[a] $N = 495$.
[b] $N = 356$.
[c] $N = 66$.

Test-retest reliabilities with about a 1-mo. interval were obtained from a group of employees ($N = 66$) of a large eastern insurance company. The coefficients for each subscale are shown in Table 2. The test–retest reliabilities are as high as could be expected, given the level of internal consistency of the subscales.

Table 3 summarizes the relationships among the subscales based on the intercorrelations of the subscale totals. Of the intercorrelations among subscale totals, 11 were significantly greater than zero ($p < .01$). The data seem consistent with the hypothesized two-factor dichotomy of work-related values (intrinsic) and reward-related values (extrinsic). The intrinsic values are Pride in Work, Activity Preference, and Job Involvement. The correlations among these variables were all substantial and positive in direction. The correlation between the extrinsic variables of Attitude toward Earnings and Social Status of Job was positive and significant. Furthermore, all of the correlations across the intrinsic–extrinsic groupings were either nonsignificant or negative. Attitude toward Earnings was negatively related to each of the intrinsic variables. The Upward Striving measure, which is both intrinsic and extrinsic in its character, showed positive correlations with both sets of scales.

Principal components analysis revealed the presence of six factors that accounted for 36% of the total variance in the matrix (see Table 4). The six factors were rotated by the quartimin method (Harman, 1967), which was selected because it places no restriction on the obliqueness of the solution. The correlations among factors were relatively low, however, ranging from − .20 to .33 with a median value of − .01. The first factor extracted represented

TABLE 3

INTERCORRELATIONS OF SUBSCALE TOTALS FOR THE SURVEY OF WORK VALUES

Subscale	Status	Activity	Striving	Earnings	Pride
Status					
Activity	.00				
Striving	.18*	.11			
Earnings	.27*	−.24*	.18*		
Pride	−.06	.60*	.14*	−.30*	
Involvement	.05	.36*	.23*	−.19*	.40*

Note.—$N = 495$.
* $p < .01$.

TABLE 4

SAMPLE ITEMS AND THEIR FACTOR LOADINGS FOR THE SIX FACTORS OF THE SURVEY OF WORK VALUES

Sample item	Factor loading
Intrinsic values	
A worker should feel some responsibility to do a decent job whether or not his supervisor is around.	.61
If the person can get away with it, he should try to work just a little slower than the boss expects him to.	−.57
The best job that a worker can get is one which permits him to do almost nothing during the working day.	−.57
Organization-Man Ethic	
A man should feel a sense of pride in his work.	.86
If a worker keeps himself busy on his job, the working day passes more quickly than if he were loafing.	.83
A worker who does a sloppy job ought to feel a little ashamed of himself.	.65
Upward Striving	
If a man likes his job, he should be satisfied with it and should not push for a promotion to another job.	−.68
A worker is better off if he is satisfied with his job and is not concerned about being promoted to another job.	−.64
A promotion to a higher-level job usually means more worries and should be avoided for that reason.	−.50
Social Status of Job	
The man who holds down a good job is the most respected man in the neighborhood.	.65
My friends would not think much of me if I did not have a good job.	.64
Having a good job makes a person worthy of praise from his friends and family.	.62
Conventional Ethic	
Doing a good job should mean as much to a worker as a good paycheck.	.64
A man should always be thinking about pulling himself up in the world and should work hard with the hope of being promoted to a higher level job.	.57
There is nothing as satisfying as doing the best job possible.	.54
Attitude toward Earnings	
A man should choose the job which pays the most.	.76
A man should take the job which offers the most overtime if the regular pay on the job is about the same.	.64
A man should choose one job over another mostly because of higher wages.	.62

an intrinsic dimension in that it was predominated by significant loadings for items from the Pride in Work, Job Involvement, and Activity Preference subscales. Separate factors were extracted for Attitude toward Earnings, Social Status of the Job, and Upward Striving. The items that loaded on a factor named Conventional Ethic characterized a high level of the work value dimension that they represented and were strongly endorsed by most *S*s. (Mean endorsement of the items was 5.16 on a 6-point scale and the mean scale value was 4.36 on a 5-point scale.) One other factor accounting for 21% of the common variance was tentatively named an Organization-Man Ethic.

DISCUSSION

Reallocation resulted in six discriminably different sets of items for which a considerable degree of rater agreement was obtained. Despite the apparent success of the reallocation procedure, the internal consistencies of the SWV subscales were not as high as one would prefer. The coefficient alpha reliabilities, which were mainly in the .60s, are about what one would expect, considering the range of scale values within each subscale and the small number of items within each subscale. Since approximately three times this number of items is usually required to achieve a reliability coefficient of .80, extreme dissatisfaction with the results of the analysis is not warranted. Yet, in light of the extensive prescreening of items by allocation, one may wonder why the homogeneity of the subscales was not greater.

Two possible explanations for the generally low item intercorrelations are worth mentioning. One such possibility is that the marginals for computation of interitem correlations are not comparable, due to the deliberate spread of items along the continuum, rather than close to the 50% endorsement level. We decided on the strategy for a wide range with the expectation that the 50% point would shift widely among samples with different background or situations, as in comparisons of social classes, or ethnic groups. This strategy resembles that of Guttman (1947) in seeking reproducibility rather than high interitem correlations.

Another possible reason for the low internal consistencies may be that the subscales of the SWV are factorially complex. In view of the magnitude of the item-total correlations, however, more than just a few strong factors would have to be demonstrated if the internal consistencies are to be ascribed to factorial complexity (Nunnally, 1967). The factor analysis did indicate, in part, a lack of correspondence with the results of the reallocation procedure in terms of discrimination of the intrinsic subscales. It did not reveal the presence of many strong factors to account for the relatively low internal consistencies and item-total correlations. The question raised by the factor analysis was whether to use the subscales determined by reallocation or those determined by the factor analysis. The lack of correspondence between the two procedures was not surprising; for although both factor analysis and reallocation are suitable methods for obtaining subscales, they are not alternate methods for achieving the same ends. Factor analysis clusters items in terms of common response patterns, whereas reallocation groups items in terms of whether they have been judged to be relevant to a particular defined category. The relative quality of prediction and meaningfulness of the relationships of subscale scores with relevant variables should indicate which method of defining subscales is preferable. The authors, therefore, elected to use the subscales determined by both methods in subsequent studies, and to delay deciding in favor of one over the other set of subscales.

A problem often associated with scales of this type is that of social desirability. While a social desirability effect is usually regarded as objectionable and as a source of bias, there is reason to doubt whether this criticism is appropriate for the measurement of values. If one is concerned about the extent to which a group of individuals shares a common value pattern, then in a very real sense it is social desirability that is being measured. Of course, measured differences in work values between diverse cultural groups are necessary if the SWV is to have value as a research instrument.

Some Applications of the Survey of Work Values

Occupational Level

The SWV was used to determine whether the subscales could discriminate between occupational groups. Wollack (1968) performed a

TABLE 5

UNSCALED VECTORS

Subscale	I	II
Status	.281	.516
Activity	.726	−.341
Striving	−.071	.331
Earnings	−.528	−.335
Pride	−.298	−.140
Involvement	−.144	−.613

TABLE 6

SCALED MEAN SCORES ON DISCRIMINANT FUNCTIONS

Group	I	II
Professional	1.54	−6.76
Supervision	1.34	−6.84
Skilled-technical	.88	−6.80
Clerical	1.33	−7.39
Semiskilled-unskilled	.51	−7.04

discriminant analysis using five occupational groups ($N = 449$) ranging from unskilled employees through professionals.

Two statistically significant discriminant functions ($p < .05$) were extracted with latent roots of .14 and .06, respectively (chi-square test for the dimensionality of discriminant space). Latent vectors for Functions I and II are given in Table 5. Discriminant scores were computed for each person in the sample. The procedure for obtaining a person's scores on the two discriminant functions consisted of weighting each of his subscale totals by the corresponding scaled vectors and summing these products over subscales. Group means are presented in Table 6.

The first discriminant function, which accounted for approximately 61% of the between-group variance, contrasted Activity Preference (positively weighted) with Attitude toward Earnings (negatively weighted). Table 6 suggests a blue-collar—white-collar continuum.

One word of caution should be raised regarding these data. One might be tempted to conclude without reservation that white-collar workers have a more favorable view of work-related activity than do blue-collar workers. However, activity probably means different things to different groups of workers. There is nothing within the items of the SWV that specifies the precise nature of job activity to which this subscale refers. To blue-collar workers, activity may imply physical activity, while to white-collar workers it may imply sedentary activity. Although this observation does not diminish the meaningfulness of these findings, it does suggest the need for cautious interpretation.

The second discriminant function, which accounted for about 29% of the between-group variance, was characterized primarily by a high positive weight for Social Status of Job and a high negative weight for Job Involvement. This function seems to describe a dimension of clerical versus other workers. Group differences appear to be small among professionals, supervisors, skilled-technical workers, and semiskilled-unskilled workers. Clerical workers, however, stood lowest on this discriminant function and were relatively distant from the other occupational groups.

Biographical Factors

The relationship between biographical factors and the subscales of the SWV has also been examined. Wijting (1969) studied the relationship of background variates and work values among unemployed persons ($N = 133$). A canonical regression analysis revealed a relationship ($R_c = .51$, $p < .05$) between a criterion composite of Activity Preference (− .71),[4] Job Involvement (− .58), and Social Status of Job (.43), and a predictor composite including sex (− .51), race (− .28), education (− .45), urban versus rural background (− .28), parents' education (.31), and amount of police trouble (.38). The criterion composite reflected an emphasis on the social rewards of being employed but a devaluation of activity and involvement on the job. The corresponding predictor composite described a white, rural, poor male whose family has had some police trouble and who has had less education than his parents.

Goodale (1969a) performed a similar analysis but with different predictors and civil service employees ($N = 342$). This study analyzed subscales determined by both methods. For the allocated subscales, Attitude toward Earnings (.75), Activity Preference (− .56), and Social Status of Job (− .43), were related ($R_c = .51, p < .001$) to race (− .65), occupational level (− .60), area of the country from which the person came (− .55), and supervisory level (.42). The combination of predictors, describing a negro whose parents came from the South and who held a low-level, nonsupervisory job, was associated with high attitude toward earnings, low preference for activity at work, and low social status of being employed. In the analysis of factored subscales, Attitude toward Earnings (.73), Organization-Man Ethic (− .54), and Conventional Ethic (− .48) were related ($R_c = .50, p < .001$) to race (− .65), area of the country from which the person came (− .48), occupational level (− .37), marital status (− .29), and supervisory level (.35). This pair of composites was characteristic of a married Negro; raised in the South; occupying a low-level, nonsupervisory job; who placed great value on earnings but did not subscribe to the Organization-Man Ethic or the Conventional Ethic.

---

[4] All loadings used to interpret the vectors are correlations of the variable with its composite rather than canonical regression weights.

SUMMARY

The SWV has met some of the common criteria (Nunnally, 1967) for construct validity. First, the success of the reallocation procedure demonstrated that the six work values are discriminably different from one another and that the items well represent the constructs that they were intended to measure. Second, the internal consistencies of the subscales are relatively high in view of the small number of items comprising each subscale. Third, SWV scores have discriminated meaningfully among occupational groups and have correlated substantially with background variables that have been associated with other measures of work values. The SWV is, in its present format, considered to be a useful research tool. Work is currently under way to improve the scale and to investigate new scoring techniques.

REFERENCES

BLOOD, M. R. Work values and job satisfaction. *Journal of Applied Psychology*, 1969, **53**, 456–459.

GOODALE, J. G. A canonical regression analysis of work values and biographical information of civil service employees. Unpublished manuscript, Bowling Green State University, 1969. (a)

GOODALE, J. G. Principal components analysis of the Survey of Work Values. Unpublished manuscript, Bowling Green State University, 1969. (b)

GUTTMAN, L. A. The Cornell technique for scale and intensity analysis. *Educational and Psychological Measurement*, 1947, **7**, 247–280.

HARMAN, H. H. *Modern factor analysis.* Chicago: University of Chicago Press, 1967.

KILPATRICK, F. P., CUMMINGS, M. C., JR., & JENNINGS, M. K. *The image of the federal service.* Washington, D. C.: The Brookings Institute, 1964.

NUNNALLY, J. C. *Psychometric theory.* New York: McGraw-Hill, 1967.

ROSENBERG, M. *Occupations and values.* Glencoe, Ill.: The Free Press, 1957.

SMITH, P. C., & KENDALL, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 1963, **47**, 149–155.

STEFFLRE, B. Concurrent validity of the Vocational Values Inventory. *Journal of Educational Research*, 1959, **52**, 339–341.

SUPER, D. E. *The psychology of careers.* New York: Harper, 1957.

SUPER, D. E. The structure of work values in relation to status, achievement, interests, and adjustment. *Journal of Applied Psychology*, 1962, **46**, 231–239.

WEBER, M. *The Protestant Ethic and the spirit of capitalism.* New York: Scribner, 1958.

WIJTING, J. P. A canonical regression of background factors and work values among underprivileged workers. Unpublished manuscript, Bowling Green State University, 1969.

WOLLACK, S. The work values of occupational groups: A multiple discriminant analysis. Unpublished manuscript, Bowling Green State University, 1968.

(Received May 21, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 339–342

ENVIRONMENTAL AMBIGUITY AND LOCUS OF CONTROL AS INTERACTIVE INFLUENCES ON SATISFACTION

ABRAHAM K. KORMAN[1]

New York University

This study hypothesized that environmental ambiguity is negatively correlated to satisfaction only for those with low control over their behavior. The hypothesis was supported in one of three studies, with a major finding being the unimportance of overall. An overall homeostatic explanation is offered in that it is proposed that situations of high ambiguity with their implied high reliance on people rather than rules and regulations are not dissatisfying if they are in balance and congruent with one's perceptions and beliefs about the nature of the world.

There are at least two major threads in the area of environmental correlates of satisfaction which seem to be commonly found. The first of these is the finding that environmental ambiguity is negatively correlated with satisfaction with that environment (cf. Kahn, Wolf, Quinn, Snoel, & Rosenthal, 1964). That is, to the extent that environments do not provide consistent guides to behavior and that whatever guides that are provided are (*a*) constantly changing; and (*b*) may, on occasion, be contradictory; then, to that extent, it is often found that dissatisfaction is the result. Second, there is, perhaps, an even more common finding that satisfaction is positively correlated with the extent to which that environment provides the opportunity to behave in a self-controlled, independent, nonsubmissive manner. In other words, the extent to which an environment is ego-supportive in that it enhances self-esteem by allowing one to behave in a competent, internally controlled manner is highly correlated with satisfaction with that environment (Tiffin & McCormick, 1965).

It would appear, then, from the viewpoint of both the organization theorist and the organizational administrator, that the way in which satisfaction with the organizational environment could be maximized is relatively straightforward. Briefly, it can easily be seen that the minimizing of ambiguity and the maximizing of self-directedness should result in the utmost satisfaction, whether the satisfactions we are talking about are the job satisfaction of the employed individual, the satisfactions of the university student, or similar situations.

Unfortunately, while such a solution seems to be theoretically adequate, it is, increasingly, in our society becoming extremely difficult to achieve. The reason for this is that ambiguity, change, and the like are becoming an inevitable part of organizational life. Thus, we live in a dynamic society, where values, norms, consumer tastes, and the like, are constantly changing and constantly ambiguous and where the administrator himself must keep introducing changes into the organization itself if it (i.e., the organization) is to remain viable and healthy in this dynamic world. Hence, the attainment of role-prescription clarity and lack of ambiguity, while desirable in order to gain a satisfied organizational membership, is not one that is easily attainable in practice.

The question, therefore, arises as to how can an organization maximize satisfactions among its members? In this article we will test the hypothesis that the significance of environmental ambiguity for satisfaction varies as a function of the degree to which the environment provides a sense of high self-esteem to the individuals in the environment. To the extent that the latter is the case, we would predict that people in high self-control environments would develop a more optimistic opinion of the capabilities of both oneself and of different others. Such an increase in the perceived competency of oneself and of different others would tend to lead, the present author believes, to (*a*) a greater confidence in one's abilities to handle the change situation where rules for behavior will become less salient and less useful; and (*b*) a greater tolerance and acceptance of the possible changes in human relationships implicit in the abstract concepts of difference, newness, and potential difference (or change) in general. Hence, the growth of such potential-difference situations as change and role ambiguity (i.e., when nothing is clear and new guidelines are being developed) would not be as likely to lead to uneasiness and dissatisfaction as if the person had been in a low self-control (or high external control) situation, where his learning about the characteristics of himself and others would more likely have been of a negative nature.

---

[1] Requests for reprints should be sent to Abraham Korman, Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.

Our specific prediction, then, is that the relationship between environmental ambiguity and satisfaction should be lower, or not exist at all, for people in high self-control environments as opposed to those situations marked by low self-control. Should this hypothesis be supported, it would be legitimate to infer that the changing, dynamic characteristics of modern society and the constantly increasing administrator-induced changes do not necessarily have to be causes of low satisfaction. Rather, such effects could be modifiable by first developing the high self-control environment.

METHOD AND PROCEDURE

Instrument. A Likert-type questionnaire was developed that was designed to measure college students' perceptions of their university environments as to the degree to which they saw it as being (*a*) ambiguous and conflicting in terms of behavioral and attitudinal prescriptions, and (*b*) ego-enhancing in terms of the degree of self-control of behavior allowed and level of competency expected. Examples of the types of statements that concerned themselves with perceptions of ambiguity were "the professors at this school are generally pretty similar to one another in basic values"; "the information you need to survive around here changes from week to week"; and "frequently, students around here are expected to achieve objectives and goals which conflict with one another and are mutually contradictory." In terms of the degree of self-control expected, samples of the types of statements used are "the general policy here seems to be that you should take as much, if not more, responsibility in planning your work as the faculty"; "the administration here tends to adopt the approach that it is only their opinions which count and not yours"; and "the professors at this college believe that they do not have to continually pressure you and watch you in order to insure that you keep your mind on your studies." There were a total of 19 conflict and 10 self-control items, all of them answerable by an agree-disagree format, interspersed with 14 buffer items and 4 general overall "satisfaction with college or university" items. These last items were highly correlated with one another (average item inter-*r* for a sample of 50 was .72), and hence were summed to serve as the measure of satisfaction in this study.

While test-retest reliabilities for the entire scale are not known, the intrascale reliabilities for each area seem to be sufficiently high. In addition to the average item inter-*r* mentioned above for the satisfaction area, odd-even reliabilities for the conflict and control area were .84 and .81, respectively, when corrected by the Spearman-Brown formula. In addition to these reliabilities, some evidence supporting the construct validity of the scales did become available during the course of the study, as will be indicated below.

Subjects. Three separate studies were undertaken of three different universities in the New York metropolitan area. However, since the design of each of these was the same, varying only in sample, the research procedure can best be thought of as an original study and two separate replications. The first university studied was a large private institution located in the heart of the city, the second was also a private institution but located in a near suburb, while the third was a state institution located in a far suburb distinctly more rural in character. In addition to these differences, the first and third schools also differed from the second in that they had been the scene of considerable strife and changing norms and values, whereas the second had been relatively placid, or at least placid enough not to hit the newspapers.

The sample sizes were 202 for Study 1, 90 for Study 2, and 91 for Study 3. All questionnaires were distributed toward the close of the 1968–1969 academic year with the samples in all cases consisting of undergraduates recruited from dormitories, psychology classes, and various other sources. Thus, while in no case could the sample be considered a random one, a wide variety of sources were utilized, suggesting some representativeness.

Analysis. For each study, attempted median splits on both variables were developed and then *t* tests for differences in satisfaction were computed for high and low ambiguity at each level of self-control. While the splits were reasonably good for Studies 2 or 3, they were not as good for Study 1. However, reasonably sized cells were obtained there also. In addition control checks were made to be certain that there were no significant differences in ambiguity at each level for each level of self-control (i.e., high ambiguity was at the same level for both levels of self-control, etc.). In each case, there were no significant differences, thus indicating sufficient commonality of ambiguity at each level for different levels of self-control.

RESULTS

The results of the research are summarized in Table 1, indicating support for the hypotheses only in Study 2.

In the other two studies the most consistent result was the lack of importance overall for ambiguity as a determinant of satisfaction. More specifically, in the case of the high self-control environments, where we did not predict any satisfaction differences as a function of environmental ambiguity, this prediction was supported in all three instances. On the other hand, the findings for the low self-control environment, where variations in satisfaction as a function of environmental ambiguity were predicted, were not supportive, except for Study 2. In the other two studies, ambiguity was no more related to satisfaction for the low groups than for the high. Briefly, ambiguity seems to be far less important as a determinant of satisfaction than are the control aspects of the environment.

TABLE 1

MEAN SCORES ON SATISFACTION AS A FUNCTION OF ENVIRONMENTAL AMBIGUITY AND SELF-CONTROL

Group	$\bar{X}$	*SD*	*t*
Study 1			
High ambiguity & high self-control^{a}	10.45	2.6	.79
Low ambiguity & high self-control^{b}	10.91	2.4	
High ambiguity & low self-control^{c}	9.16	2.8	.01
Low ambiguity & low self-control^{d}	9.20	3.0	
Study 2			
High ambiguity & high self-control^{e}	9.84	2.3	1.35
Low ambiguity & high self-control^{f}	10.64	2.2	
High ambiguity & low self-control^{g}	8.19	2.2	2.55*
Low ambiguity & low self-control^{h}	9.90	2.0	
Study 3			
High ambiguity & high self-control^{i}	11.70	2.3	.73
Low ambiguity & high self-control^{j}	10.98	2.4	
High ambiguity & low self-control^{j}	8.94	2.7	.30
Low ambiguity & low self-control^{k}	9.17	2.1	

[a] $n = 64$.
[b] $n = 58$.
[c] $n = 59$.
[d] $n = 30$.
[e] $n = 17$.
[f] $n = 31$.
[g] $n = 26$.
[h] $n = 16$.
[i] $n = 27$.
[j] $n = 18$.
[k] $n = 28$.
* $p = .01$.

DISCUSSION

Looking only at Study 2 for a moment, the results can generally be interpreted as supportive of the hypothesis which guided this research. This is a finding of considerable significance for the administrator since it is possible for him to control the self-control aspects of the organization but increasingly not the ambiguity, both within and without. Thus, the finding that self-control is more important than ambiguity means that the administrator need not be loath to introduce needed changes into the organization because of the need to maintain harmony, etc. Similarly, there is no need to demand rigid role prescriptions, as demanded by traditional organization theory, because of their stabilizing nature. The key, apparently, seems to be to increase self-confidence since, by so doing, we can increase satisfaction overall and, at the same time, reduce fear of the unknown and the change. By so doing, changes in organization may be introduced and the result will not be dissatisfaction with its frequent concomitants of turnover and withdrawal.

Perhaps an equally significant implication is raised by the lack of significance overall in Studies 1 and 3 for ambiguity as a cause of dissatisfaction in general. This, too, suggests to the administrator that he need not be afraid to either accept or introduce rule and regulation relaxation. We can only conjecture here as to possible reasons, but one possibility may be due to the activation level of the individual involved (Fiske & Maddi, 1961). Thus, the campuses of both the Study-1 and Study-3 groups had been involved in ferment for so long and changes had become so commonplace that perhaps they had become accepted as part of the landscape, even by the low self-control people. Hence, there was no sense getting very excited or unhappy about such ambiguity and lack of clarity. The implications for the administration here, provided our explanation is a valid one, are very similar to the implications we drew from the Study 2 findings, in that in both cases, change has become increasingly accepted and people are not afraid of it, whether such acceptance comes from lack of fear of difference or from generalized adaptation to a normal state of affairs (such normality being increasingly dynamic).

As support for this interpretation of our findings, we may consider some evidence supporting the construct validity of our scale. We have said that the schools involved in Studies

1 and 3 had been involved in considerable change and unrest, whereas the school in Study 2 had not and, in fact, have built our explanatory hypotheses around such differences in order to explain our results. Hence, it is supportive of the construct validity of our scale and our line of reasoning that the students involved in Studies 1 and 3 had significantly higher overall scores on perceptions of environmental ambiguity than those from Study 2. (The means and standard deviations for Studies 1, 2, and 3, respectively, were 67.8 and 9.8, 53.2 and 7.9, and 65.5 and 8.6. Studies 1 and 3 are significantly different from 2, but not from each other.)

There is, however, a question which might be raised regarding our argument. Essentially, we are suggesting a homeostatic interpretation of our findings in that we are proposing that situations of high ambiguity with its implied high reliance on people rather than specified rules and regulations are not dissatisfying if they are in balance and congruent with one's perceptions and beliefs of the world. Thus, such perceptions and beliefs might be composed of perceptions of the typical nature of things around one (i.e., Studies 1 and 3), or one's beliefs concerning the nature of people (i.e., Study 2), with the strength of each effect perhaps varying according to the specific situation. Thus, in the cases we have studied here, the degree of variability in the environment in general in Studies 1 and 3 was strong enough so that tolerance for ambiguity developed for the general sample overall and it was not necessary to change the self-control aspects of the environment in order to increase tolerance for ambiguity. (Although increasing self-control would have increased satisfaction as a general effect in all universities, this was not a matter under specific investigation here.) On the other hand, in the Study 2 situation, where the environment was in general more stable, increasing self-control would seem to be necessary to increase tolerance and acceptance of ambiguity.

Given this interpretation, the major question, then, is if people develop acceptance of ambiguity in either of these two ways, why don't they actually like them even more than the stable stituations, as a strict homeostatic prediction might imply, rather than there being no differences? One answer to this is that such acceptance and satisfaction is probably a developmental process over time, with the degree of preference for ambiguity as opposed to stability being a function of the amount of time and experience one has had in self-control and variability situations. The longer one has been in such situations and the more one has learned that this is the normal state of affairs, the more we would predict such differential preference for variability as opposed to stability. However, in this case, we were not dealing with substantive individual differences in belief systems concerning the nature of the world which are solidly based on one's life experiences. Rather, we were dealing with contemporary environmental influences, influences which in many cases might have been diametrically in opposition to a person's characteristic belief structure. Hence, it would seem to be legitimate to consider the results found here to be reasonably supportive of our interpretation.

To summarize, then, the findings of this research would seem to point to the conclusion that justification of rules and regulations as being more satisfying, due to their fulfilling needs for order and predictability, may be overstated. Apparently, ambiguity and change are not dissatisfying if the individual's belief system is first altered so that he sees such characteristics as being consistent with the nature of the world as he knows it. Finally, such cognitive change may be introduced by the administrator, but it may also occur as a function of factors beyond his control.

REFERENCES

FISKE, D., & MADDI, S. (Eds.) *Functions of varied experiences.* Homewood, Ill.: Dorsey Press, 1961.

KAHN, R. L., WOLF, D. M., QUINN, R. P., SNOEL, J. D., & ROSENTHAL, R. A. *Organizational stress: Studies in role conflict and ambiguity.* New York: Wiley, 1964.

TIFFIN, J., & MCCORMICK, E. *Industrial psychology.* (5th ed.) Englewood Cliffs, N. J.: Prentice-Hall, 1965.

(Received February 11, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 343–348

MORALE AND PRODUCTIVITY CONSEQUENCES OF GROUP LEADERSHIP STYLE, STRESS, AND TYPE OF TASK[1]

LEONARD L. ROSENBAUM[2]
Montgomery College

AND

WILLIAM B. ROSENBAUM
New York Institute of Technology

Sixty male undergraduate *S*s worked on structured and unstructured tasks under stressful or nonstressful conditions in groups of four. A confederate of *E*, who was one of the members of each four-person group, adopted either a democratic or authoritarian leadership style. Measures of performance, time to completion, and satisfaction were used in separate analyses employing a Lindquist Type VI analysis of variance. The *S*s under stress performed best under authoritarian leadership, whereas, those not under stress performed best under democratic leadership. The *S*s under authoritarian leadership conditions took significantly longer to complete their tasks. In addition, *S*s performed better, worked faster, and most enjoyed the structured task.

Since the pioneering research of Lewin, Lippitt, and White (1939) many studies have further investigated the consequences of different styles of leadership. Anderson (1963), in a review of research in the area, concluded that: "The evidence available fails to demonstrate either authoritarian or democratic leadership is consistently associated with higher productivity. In most situations, however, democratic leadership is associated with higher morale [p. 160]." Bass (1960) derived a set of propositions about the need for more directive leadership under stress and the greater acceptability of structure under these circumstances. Earlier, Fleishman, Harris, and Burtt (1955) had found that as the time schedules for production became more demanding, the rated effectiveness of foremen of those departments was increasingly correlated with leadership lower in consideration and higher in structure. In departments with less demanding time schedules, the reverse pattern tended to hold.

A theory based on psychoanalytic tenets suggested by Devereux (1955) and Alexander (1955) argues that when an *S* is confronted with an unsolvable task, he would return to dependence on external parental help rather than rely on his own ego resources. Hamblin (1958) provides some support for this theory in that his experimental *S*s who were subjected to a crisis situation in the laboratory were more subject to influence than controls. Mulder (1963) found similar results in a study involving independent grocers approximately half of whom were told that a supermarket would be opening near them. The grocers in the high-threat condition apparently showed a tendency to seek strong leadership. Polis (1964) studied Australian grade school children and found that *S*s who chose democratic types of leadership during peace time chose omnipotent forms of leadership during an imaginary crisis situation.

These results can also be explained by McGinnies (1970), who adopts a functional analytic approach and advocates the view that "a crisis establishes a deprivation state for a group, so that the consequences of effective leadership will be more reinforcing than otherwise. The circumstances are roughly analogous to those of depriving an experimental animal of food so that food becomes a more effective reinforcer [p. 238]."

Korten (1962) developed a model concerned with some of the situational factors that determine group preferences for authoritarian or democratic leadership. He suggests that authoritarian leaders maintain their power in highly stressful, highly structured goal situations, whereas, democratic leadership is sought under less stressful, poorly structured situations.

---

[1] The authors are indebted to James Brewer, Steven Futterman, Ronald Ginn, and Charles Guyker who served as either *E* or accomplice and conscientiously collected the data for this study in partial fulfillment of the requirements of an advanced experimental social psychology course at the University of Maryland. Also, the authors are grateful to Elliott McGinnies for his critical reading of the manuscript.

[2] Requests for reprints should be sent to Leonard L. Rosenbaum, Department of Psychology, Montgomery College, Rockville, Maryland 20850.

Fiedler, Meuwese, and Oonk (1961) and Fiedler (1962) studied leadership behavior in a variety of groups working on different tasks. Under a pleasant group atmosphere Fiedler found a positive correlation between the leader's esteem for Least Preferred Coworker (LPC) score and productivity. Under unpleasant group atmosphere, or when groups were heterogeneous and stress resulted, the relationship between LPC score and productivity was negative. Fiedler (1964) developed a contingency model that predicts that the effectiveness of leadership style depends on the extent to which the group's task and organization facilitates or inhibits the leader's ability to exert influence. A low-LPC leader would function best in a negative group atmosphere and on unstructured tasks. He also notes that a low-LPC leader would be effective if the task were especially structured. In the intermediate situation, however, the high-LPC leader would be most effective. In a study of officers and recruits at a Belgian naval training center Fiedler (1966) provides some confirmation of the curvilinear relations predicted by the theory.

Consequently, it was hypothesized that, under stress, groups under authoritarian leadership would perform better and be more satisfied, whereas, democratic leadership groups would perform better and be more satisfied under the nonstress condition. It was further anticipated that groups under authoritarian leadership would be more satisfied and perform better on the structured task, whereas, democratic groups would show a preference (in satisfaction and performance) for the moderately unstructured tasks.

METHOD

Subjects

Sixty male undergraduate introductory psychology students at the University of Maryland served as *S*s for this study. The Psychology Department required that all students in the introductory course earn research credits by participating as *S*s in research projects. The *S*s who volunteered to participate in this study were randomly assigned to 20 experimental groups. Each group consisted of three naïve *S*s plus a confederate of *E*.

Procedure

Under either stressful or nonstressful conditions each group of *S*s solved two structured or mechanical tasks and two unstructured or problem-solving tasks under both autocratic and democratic leadership conditions. The independent variables in this study therefore were: type of leadership, amount of stress, and type of task. The dependent variables examined were group performance, group satisfaction, and time to solution.

Leadership manipulation. A confederate of *E* participated with each group of three naïve *S*s. The confederate was an undergraduate enrolled in an advanced experimental social psychology course at the University of Maryland. He arrived at the laboratory at the scheduled hour, as did each of the three naïve *S*s. He represented himself as a student in the introductory psychology course and maintained this deception until all data for this study were collected. The confederate was trained in the leadership behaviors appropriate to each of the leadership styles identified by Lewin and Lippitt (1938). Autocratic leaders attempted to: (*a*) determine the group's policy, (*b*) go only one step at a time when giving directions so that the group was uncertain as to what to do next, (*c*) structure what each group member would do and with whom, (*d*) criticize and praise individual members of the group, and (*e*) remain aloof from the group. On the other hand, democratic leaders attempted to: (*a*) have the group determine their work policies, (*b*) explain and provide perspective of work procedures, (*c*) allow group members to choose between alternative procedures, (*d*) criticize and praise the group as a whole, and (*e*) become a group member in spirit.

It is interesting to note that ratings of liking of the confederate obtained on a questionnaire administered after each task was completed did not differ significantly ($t = .73$) from authoritarian and democratic conditions. This questionnaire also asked each student to rate his satisfaction with his performance and the group's performance on the just-completed task.

Task manipulation. Two unstructured and two structured tasks were utilized in this research. The sequence in which these tasks were performed under the different leadership conditions was counterbalanced to minimize order effects.

One structured task required *S* to place in numerical order a deck of 45 randomly arranged cards; to alphabetize 45 cards, each with a nonsense syllable on it; and then to paper clip one number card and one nonsense syllable card to each other. Finally, *S* was to sort the resulting stack of cards into odd- and even-numbered piles. The second structured task required *S* to make six sketches of each of six stimulus objects (e.g., a bus), to crayon these with prescribed colors, and to cut the 36 objects out and paste them on six sheets of paper such that six different objects appeared on each composite.

An unstructured task involved asking *S*s to classify 18 sentences from Bales' (1952) Standard Human Relations Problem, according to Bales' 12 categories for interaction process analysis. The other unstructured task required *S* to determine whether 18 syllogisms were true or false.

A maximum of 10 min. was allowed for completion of each of these tasks. Further, *S*s were told that they might work on each task in any sequence. They were informed that the experiment was primarily concerned with the productivity of the group, but each *S* was encouraged to work quickly and accurately.

Stress manipulation. At the outset of the experimental session, a group was exposed to either the high- or low-stress manipulation. The *S*s in the high-stress condition were told that the tasks they would be asked to complete were from an established test and that each group's performance would be compared to performance of other groups at the college. They were also told that *S* should indicate his name and class on all materials so that a permanent record of his scores might be kept, and a copy could be provided to his psychology instructor.

The *S*s in the low-stress condition were told that *E* just designed the tasks and hoped they were interesting. These *S*s were also told that individual responses were not important to *E*, only group productivity was. Consequently, it was not necessary for *S* to sign or identify himself on any of the various answer sheets.

As an additional check on whether *S*s, in fact, felt stress, the stress manipulations used in this study were subsequently employed with 42 undergraduate psychology students at the New York Institute of Technology. The students worked in groups of three. Twenty-one students were exposed to the high-stress manipulation. They reported significantly greater stress ($t = 2.26$; $p < .05$) than did 21 of their colleagues who were exposed to the low-stress manipulation. Stress was measured by responses to a 5-item questionnaire consisting of statements regarding the stress, tension, anxiety, concern, and pressure that they experienced because of the experiment. The *S*s indicated the extent to which they agreed or disagreed with each item, using seven graded alternatives.

RESULTS

Measures of performance, satisfaction, and time to solution were obtained for all experimental *S*s. Mean group responses for each of the dependent variables were obtained from each group of three experimental *S*s, for each level of the independent variables of leadership, task, and stress. These group means were the basic data used in each of the following three analyses. The Lindquist (1953) Type VI analysis of variance represents a three-way analysis of variance in which two of the variables are repeated measures. In this study, the stress manipulation employed necessitated obtaining measures from two groups of *S*s. The independent variables of leadership and type of task, however, represent repeated measures on the same *S*s. Consequently, a Lindquist Type VI analysis of variance was completed for each of the dependent measures of group performance, group satisfaction, and time to solution.[3]

TABLE 1

ANALYSES OF VARIANCE OF MEAN GROUP PERFORMANCE SCORES, TIME TO TASK COMPLETION, AND SATISFACTION RATINGS

Source	*df*	Performance score		Time to task completion		Satisfaction rating	
		MS	*F*	*MS*	*F*	*MS*	*F*
Between-*S*s	19	.01		2.59		11.17	
Stress (C)	1	.00	.15	5.96	2.48	6.24	.55
Error (b)	18	.01		2.40		11.44	
Within-*S*s	60	.02		1.74		9.17	
Task (A)	1	.64	67.06**	7.25	6.18*	60.64	6.69*
A × C	1	.00	.35	.02	.02	2.82	.31
Leadership (B)	1	.01	1.05	5.47	5.87*	3.33	.46
B × C	1	.06	4.43	1.43	1.54	.07	.01
A × B	1	.01	.29	.57	.21	.88	.09
A × B × C	1	.00	.13	2.75	1.02	2.80	.28
Error (w)	54	.01		1.61		8.89	
Error$_1$ (w)	18	.01		1.18		9.07	
Error$_2$ (w)	18	.01		.93		7.41	
Error$_3$ (w)	18	.02		2.71		10.18	
Total	79	.02		1.94		9.65	

* $p < .05$.
** $p < .001$.

Performance Index

The measure of performance used in this analysis was the ratio of the number of correct responses made by *S*s to the total number of correct responses possible. The possible range of scores then, is from 0 to 1.00. The higher the index, the better the group performance. Table 1 presents the results of the analysis of variance which was performed on these data. The main effect for type of task was significant at the .001 level. Examination of the means presented in Table 2 reveals that the overall mean performance on the unstructured task was .72, whereas, the mean performance on the structured task was .90. Generally, *S*s performed substantially better on the structured task.

In addition, this analysis of variance reveals a significant ($p < .05$) interaction between leadership style and stress. The cell means depicting this interaction are also presented in Table 2. Under stress, groups with authoritarian leadership had a mean performance of .83, whereas, democratic leadership groups had a mean performance score of .80. This relation-

---

[3] The computer time for this project was supported in full through the facilities of the Computer Science Center of the University of Maryland.

TABLE 2

MEAN GROUP PERFORMANCE SCORES, TIME TO COMPLETION, AND SATISFACTION RATINGS

Task	Stress			Nonstress			Combined stress		
	Authoritarian leadership	Democratic leadership	Combined leadership	Authoritarian leadership	Democratic leadership	Combined leadership	Authoritarian leadership	Democratic leadership	Combined leadership
Group $\bar{X}$ performance index									
Unstructured	.72	.72	.72	.68	.76	.72	.70	.74	.72
Structured	.93	.88	.91	.85	.92	.89	.89	.90	.90
Combined	.83	.80	.81	.76	.84	.80	.80	.82	.81
Group $\bar{X}$ time to completion									
Unstructured	9.67	9.62	9.65	9.80	8.47	9.14	9.74	9.05	9.35
Structured	9.31	8.85	9.08	8.63	8.38	8.50	8.97	8.61	8.83
Combined	9.49	9.24	9.36	9.21	8.42	8.82	9.39	8.79	9.09
Group $\bar{X}$ satisfaction score									
Unstructured	15.80	16.43	16.12	15.30	15.07	15.18	15.55	15.75	15.65
Structured	17.33	17.63	17.48	16.83	17.77	17.30	17.08	17.70	17.39
Combined	16.57	17.03	16.80	16.07	16.42	16.24	16.32	16.73	16.52

ship was reversed for groups not exposed to the stress manipulation. For these groups, democratic leadership resulted in a mean performance of .84, whereas, autocratic leadership resulted in the lowest mean performance score of .76. As anticipated, *S*s in the stress conditions performed best under authoritarian leadership. On the other hand, *S*s under nonstress conditions tended to perform best under democratic leadership.

Time to Completion

The amount of time it took to complete each task was recorded in minutes and was used as the dependent variable in a second analysis of variance. It should be recalled that *S*s were allowed a maximum of 10 min. to complete each task. The analysis of variance of the mean group time to completion data, also depicted in Table 1, resulted in statistically significant ($p < .05$) leadership style and type of task main effects.

Examination of Table 2 reveals that *S*s, on the average, were able to complete the structured task significantly faster (8.83 min.) than the unstructured task (9.35 min.). The *S*s performed better and faster on the structured task; it was apparently easier.

Furthermore, under democratic leadership conditions *S*s needed an average of 8.79 min. to complete each task, whereas, under authoritarian leadership the significantly longer time of 9.39 min. was necessary to complete the task. In the previous analysis, performance under democratic conditions was slightly better than under authoritarian conditions, however, this effect was not significant. It appears then that the authoritarian leadership did substantially increase the amount of time needed to complete the task but did not interfere dramatically with the quality of work performed.

Satisfaction

Three 7-point Likert-type rating scales designed to measure *S* satisfaction with performance on the experimental tasks were administered after each task was completed. Mean satisfaction scores for each *S* for each type of task were computed. In turn, these mean

S-satisfaction scores were used to calculate group mean satisfaction scores to each type of task under each of the leadership conditions. The possible range of these group mean satisfaction scores was from 3 to 21. The larger the satisfaction score obtained, the greater the satisfaction. These scores represent the data submitted to the Lindquist Type VI analysis of variance. The results of that analysis also appear in Table 1. Surprisingly, virtually no difference in satisfaction under the different leadership styles is evident. The main effect for type of task was significant at the .05 level. The mean *S* satisfaction with the structured task was 17.39, whereas, for the unstructured task it was 15.65—significantly less. Apparently then, *S*s performed better, worked faster, and most enjoyed the structured task.

DISCUSSION

Although there have been conflicting reports in the literature about the relative merits of the different leadership styles on productivity, our results support the various approaches that argue that leadership effectiveness is situationally determined. Under stress, the present authors found greater productivity for autocratically led groups, whereas, in the absence of stress, democratic leadership was more productive.

In addition, the authors hypothesized that satisfaction ratings should parallel the productivity ratings obtained. This hypothesis was not confirmed. Previous evidence has, as Anderson (1963) indicates, consistently demonstrated an association of higher morale with democratic leadership. The *S*s in our study would normally be expected to prefer democratic leadership. Probably significant in this regard is the especially poor performance obtained from those *S*s who were autocratically led but were not in the stress condition. McGinnies (1970) argues that effective leadership during crises is reinforcing. Perhaps the greater satisfaction generated for autocratically led groups under stress, as hypothesized by McGinnies, in combination with the greater satisfaction of democratically led groups, regardless of which stress condition, combined to eliminate the anticipated differences.

Real differences existed between the types of tasks employed in this study. The task variable, however, did not interact with leadership style or stress. The tasks identified as unstructured probably were not perceived as such by *S*s. To college students, these tasks may have represented an intellectual sorting task not sufficiently dissimilar to the structured mechanical sorting tasks employed here.

Shaw (1962) outlined four aspects for classifying group tasks, namely: (*a*) decision verifiability, (*b*) goal clarity, (*c*) goal path multiplicity, and (*d*) solution specificity. When our tasks are evaluated along these dimensions, it appears that all tasks, in fact, were quite structured and differed primarily in the extent to which they relied on manual or verbal solutions. The better performance and shorter work time found for the structured or mechanical tasks reflected intrinsic differences between the two sets of tasks. The greater satisfaction obtained to the structured task is probably a reflection of the generalization of reinforcement effects resulting from doing better on the structured tasks. The finding of better performance under autocratic leadership across all tasks partially confirms our initial hypothesis if it is agreed that both sets of tasks are rather structured.

REFERENCES

ALEXANDER, F. On the psychodynamics of regressive phenomena in panic states. *Psychoanalysis and the Social Sciences*, 1955, 4, 104–110.

ANDERSON, R. C. Learning in discussions: A resume of the authoritarian-democratic studies. In W. W. Charters, Jr. & N. L. Gage (Eds.), *Readings in the social psychology of education*. Boston: Allyn & Bacon, 1963.

BALES, R. F. Some uniformities of behavior in small social systems. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. (2nd ed.) New York: Holt, 1952.

BASS, B. *Leadership, psychology and organizational behavior*. New York: Harper & Row, 1960.

DEVEREUX, G. Charismatic leadership and crisis. *Pyschoanalysis and the Social Sciences*, 1955, 4, 145–157.

FIEDLER, F. E. Leader attitudes, group climate, and group creativity. *Journal of Abnormal and Social Psychology*, 1962, 65, 308–318.

FIEDLER, F. E. A contingency model of leadership effectiveness: In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 1. New York: Academic Press, 1964.

FIEDLER, F. E. The effect of leadership and cultural heterogeneity on group performance: A test of the contingency model. *Journal of Experimental Social Psychology*, 1966, 2, 237–264.

IEDLER, F. E., MEUWESE, W., & OONK, S. Performance of laboratory tasks requiring group creativity. *Acta Psychologica*, 1961, **18**, 100–119.

LEISHMAN, E. A., HARRIS, E. F., & BURTT, H. E. *Leadership and supervision in industry*. (Research Monograph No. 33.) Columbus, Ohio: Ohio State University, Bureau of Educational Research, 1955.

HAMBLIN, R. L. Leadership and crisis. *Sociometry*, 1958, **21**, 322–335.

KORTEN, D. C. Situational determinants of leadership structure. *The Journal of Conflict Resolution*, 1962, **6**, 222–235.

LEWIN, K., & LIPPITT, R. An experimental approach to the study of autocracy and democracy: A preliminary note. *Sociometry*, 1938, **1**, 292–300.

LEWIN, K., LIPPITT, R., & WHITE, R. K. Patterns of aggressive behavior in experimentally created "social climates." *Journal of Social Psychology*, 1939, **10**, 271–299.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

MCGINNIES, E. *Social behavior: A functional analysis*. Boston: Houghton Mifflin, 1970.

MULDER, M. Threat, attraction to the group and need for strong leadership, a laboratory experiment in a natural setting. *Human Relations*, 1963, **16**(4), 317–334.

POLIS, T. A note on crisis and leadership. *Australian Journal of Psychology*, 1964, **16**(1), 57–61.

SHAW, M. E. *Annual Technical Report, 1962*. Gainesville, Florida, University of Florida, 1962. (Also in H. Proshansky and B. Seidenberg (Eds.), *Basic studies of social psychology*. New York: Holt, Rinehart & Winston, 1965.)

(Received April 13, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 349–352

FAVORABLE SELF-PERCEPTION, PERCEIVED SUPERVISORY STYLE, AND JOB SATISFACTION

DUANE E. THOMPSON[1]

College of Business Administration, University of Iowa

Interrelationships of favorable self-perception as measured by biographical data, perceived supervisory style of the boss, and job satisfaction of 128 administrators and professional employees of a state department of public instruction were compared using a two-factor analysis of variance design. The *S*s with a high level of favorable self-perception were less likely to perceive the supervisory style of their boss as supportive and reported lower levels of job satisfaction than those with a low level of favorable self-perception. More supportive styles of supervision were found to be associated with higher levels of job satisfaction, although the effect was moderated by favorable self-perception. Results are discussed in terms of the relationship between expectations, experience, and attitudes.

There is evidence that within the work situation, the supervisory style of the boss plays an important role in providing the opportunity for the individual to experience satisfaction of personal needs. Likert (1961, 1967), for example, describes a supervisory style that, as a part of a larger system of management, reportedly leads to higher levels of job satisfaction and performance than other, contrasting styles. This style is governed in part by the "principle of supportive relationships" that states in essence that within an organization all relationships and interactions must be such that each individual will view these experiences as supportive (i.e., serving to build and maintain his sense of personal worth). The expectations of the individual as determined by his background, values, and experience form an integral part of this principle. "If experience falls short of expectation, unfavorable attitudes occur. Where our experience is better than our expectations we tend to have favorable attitudes [Likert, 1961, pp. 1–2]."

A perceived supervisory style similar to that promoted by Likert (1961) was found by Myers (1966) to lead to high levels of job satisfaction among managers of a large electronics company. Myers did not, however, investigate the effect individual differences in expectations have on this relationship. The present study was designed to replicate Myers' study in a dissimilar organization and to explore the relationship of the individual's background of experience and expectations to his perception of supervisory support and job satisfaction. In this regard, it was initially assumed that an individual's background of experience and appraisal of abilities would combine to form a measure of his self-perception and, hence, an indication of the types of treatment the individual would expect in the superior-subordinate relationship. Specifically, it was hypothesized that *S*s with high self-perception would experience less job satisfaction from a given supervisory style than those with low self-perception.

METHOD

Subjects and data collection. The *S*s were administrators and professional employees of a state department of public instruction. Survey forms consisting of a cover letter guaranteeing anonymity, general information questions, and two separate questionnaires were completed by 121 *S*s during a bimonthly meeting of the department. Additional survey forms and self-addressed envelopes were left at the department with the understanding that the assistant superintendents would distribute the forms to those absent from the meeting. In this manner an additional seven surveys were completed for a total of 128 *S*s. Of these, 21 had been with the department less than 1 yr., 62 from 1 to 5 yr., and the remaining 45 over 5 yr. All *S*s except eight were male.

Questionnaires. Job satisfaction and supervisory style of the boss were measured with a modification of the Management Attitude Questionnaire developed by Myers (1966). The revised form, entitled Administrative Attitude Questionnaire (AAQ), was constructed with only minor word changes from the Management Attitude Questionnaire. The resulting 96-item AAQ consisted of 10 multiple-choice, 62 rating-type, and 24 forced-choice items. Principal component analysis and factor matching by means of the salient variable similarity index (Cattell & Baggaley, 1960) indicated that the factor structures of the two forms were sufficiently similar to justify the use of subscales identified by Myers.

---

[1] Requests for reprints should be sent to Duane E. Thompson, College of Business Administration, University of Iowa, Phillips Hall, Iowa City, Iowa 52240.

Individual differences information regarding *S*s' expectations, backgrounds, and experiences was collected by means of a specially constructed group of 41 multiple-choice items entitled Biographical Information Blank (BIB). Items were selected from those previously used by Morrison (1962), Gilmer (1963), and Thomson (1964). The criteria for inclusion of items were: (*a*) items should load on Morrison's scale of Favorable Self-Perception, or on the corresponding scale identified by Gilmer; and (*b*) other items from the three studies should be used only if they appeared to bear a relationship to the problem at hand. Preliminary component analysis of the BIB data yielded three factors, one of which was found by the salient variable similarity index to be statistically related to Morrison's Favorable Self-Perception scale. The other two factors could not be given a meaningful interpretation (Thompson, 1969).

Specific AAQ and BIB subscales used in this research were:

AAQ

Job Satisfaction: Items with high loadings on this subscale (previously labeled Motivation on the Job by Myers) dealt with feelings of utilizing potential, being highly stimulated, the opportunity for growth, self-esteem, and achievement on the job.

Supervisory Style of the Boss: This subscale was comprised of items that "describe the boss' style in terms of his ability to stimulate enthusiasm, the level and consistency of his expectations, his recognition of performance, his accessibility and willingness to listen to new ideas, his practices related to dispensing company information, his attitude toward risk taking and mistakes, and his sensitivity to the feelings of others [Myers, 1966, p. 59]."

BIB

Favorable Self-Perception: Items loading on this subscale dealt with level of income, type of work preferred, assessment of potential, level of education attained, value gained from education, and membership in professional organizations.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR THE AAQ AND BIB SUBSCALES

Subscale	$\bar{X}$	*SD*
Supervisory Style of the Boss	9.14	3.50
Job Satisfaction	13.04	3.81
Favorable Self-Perception	8.81	1.24

Note.—Abbreviations: AAQ = Administrative Attitude Questionnaire, BIB = Biographical Information Blank.

TABLE 2

CROSS CLASSIFICATION OF *S*S ON THE BASIS OF SUPERVISORY STYLE OF THE BOSS AND FAVORABLE SELF-PERCEPTION

Favorable Self-Perception	Supervisory Style of the Boss			Total
	Supportive	Traditional	Nonsupportive	
High	7	14	22	43
Medium	18	10	15	43
Low	17	19	6	42
Total	42	43	43	128

Note.—$N = 128$.

Procedure. Subscale scores for the two AAQ subscales and the BIB subscale were computed by assigning a value of 1 to the most favorable response option of each item and 0 to the least favorable option. The value assigned intermediate responses depended only on the number of options. If *S*s had failed to respond to an item, the mean value of the item was assigned. Means and standard deviations for the scales thus computed are presented in Table 1.

The relationship between Favorable Self-Perception, Supervisory Style of the Boss, and Job Satisfaction subscale scores was tested with a 3 × 3 factorial analysis of variance. The *S*s were divided into three groups on the basis of their score on the Supervisory Style of the Boss subscale to form column classifications for the design. The top one-third were considered as reporting to supervisors perceived as "supportive," in that favorable responses to the items of the present scale described "supportive" as defined by Likert (1961). The middle third were considered as reporting to those perceived as "traditional," a style characterized as avoiding either extreme, and the bottom one-third as reporting to bosses perceived as "nonsupportive." Row classifications were similarly devised by sorting *S*s on the basis of their Favorable Self-Perception subscale scores. Groups were simply labeled as "high," "medium," and "low" to designate the third of the distribution in which their score fell. Job Satisfaction subscale scores were the criterion measures. Unequal cell frequencies associated with this method of classification were handled by treating the univariate analysis of variance as a special case of a multivariate analysis of variance procedure that permitted unequal cell frequencies (Clyde, Cramer, & Sherrin, 1969).

RESULTS

The initial cross classification of *S*s based on subscale scores for Supervisory Style of the Boss and Favorable Self-Perception are depicted in Table 2.

As may be noted in this table, only 7 of the 43 *S*s in the high Favorable Self-Perception subgroup perceived their boss as displaying what has been labeled as a supportive style of supervision, whereas, 22 of this same group perceived their boss as exhibiting a nonsupportive style of supervision. This is in distinct contrast with the 42 *S*s in the low Favorable Self-Perception group. Of these, 17 perceived their boss as supportive, but only 6 perceived his behavior as nonsupportive. The statistical significance of this relationship is born out by an r of $-.30$ ($p < .01$) between Supervisory Style of the Boss and Favorable Self-Perception subscale scores. Job Satisfaction subscale means corresponding to the data of Table 2 are presented in Table 3.

TABLE 3

AVERAGE JOB SATISFACTION SUBSCALE SCORES FOR *S*s CLASSIFIED ON THE BASIS OF SUPERVISORY STYLE OF THE BOSS AND FAVORABLE SELF-PERCEPTION

Favorable Self-Perception	Supervisory Style of the Boss			Row $\bar{X}$
	Supportive	Traditional	Nonsupportive	
High	13.5	12.9	10.2	11.6
Medium	15.3	12.0	10.8	12.9
Low	14.7	16.0	10.0	14.6
Column $\bar{X}$	14.8	14.0	10.4	13.0

Note.—$N = 128$.

As indicated in Table 3, mean Job Satisfaction subscale scores decreased as Favorable Self-Perception increased and increased as Supervisory Style of the Boss became supportive. Highest Job Satisfaction cell means were associated with a supportive style of supervision for those classified medium or low on Favorable Self-Perception and with a traditional style of supervision for those classified as low in Favorable Self-Perception. Low Job Satisfaction cell means were associated with all levels of Favorable Self-Perception for *S*s perceiving their boss as nonsupportive.

Results of the two-factor analysis of variance performed with these data are presented in Table 4. From these results it was concluded that Supervisory Style of the Boss and Favorable Self-Perception subscale scores jointly acted in the determination of Job Satisfaction subscale scores.

DISCUSSION

By and large, Myers' (1966) findings that a supportive style of supervision was associated with higher levels of job satisfaction have been supported by this study conducted in a dissimilar organization and environment. In addition, Favorable Self-Perception, as measured by personal history data, was found to act on this relationship in two ways. First, those scoring high on Favorable Self-Perception were less likely to perceive the behavior of their supervisor as supportive than those scoring lower on the Favorable Self-Perception scale; and second, it was found that those scoring high on the Favorable Self-Perception subscale generally scored lower on Job Satisfaction.

The first question to be considered is how one can account for these results. The observed relationship between Favorable Self-Perception, Supervisory Style of the Boss, and Job Satisfaction subscale scores may be cast in terms of the theoretical framework of Likert (1961) if the following assumptions are made: (*a*) Job Satisfaction subscale scores are a measure of the subordinate's attitudes; (*b*) Supervisory Style of the Boss subscale scores form an index of the subordinate's perception of at least one category of experience in the organization; and (*c*) Favorable Self-Perception subscale scores may be interpreted as an indication of the expectations the subordinate has of his supervisor (a reasonable assumption based on the relationship with perception of supervisory style).

In this context, it would appear that for those perceiving the Supervisory Style of the

TABLE 4

SUMMARY OF TWO-FACTOR ANALYSIS OF VARIANCE: SUPERVISORY STYLE OF THE BOSS AND FAVORABLE SELF-PERCEPTION WITH JOB SATISFACTION AS CRITERIA

Source	*df*	*MS*	*F*
Supervisory Style of the Boss (Columns)	2	167.4	16.2**
Favorable Self-Perception (Rows)	2	94.8	9.2**
Rows × Columns	4	24.9	2.4*
Within	119	10.4	

Note.—$N = 128$.
* $p < .053$.
** $p < .001$.

Boss as nonsupportive, this experience was less than their expectation, hence, the uniformly low Job Satisfaction subscale mean. On the other hand, the perception of a traditional supervisory style appeared to have been sufficient to exceed the expectations of those classified as low in terms of Favorable Self-Perception, whereas, a supportive style was required to exceed the expectations of those with a medium level of Favorable Self-Perception. As indicated by the Job Satisfaction mean for those classified as low Favorable Self-Perception and reporting to a boss perceived as supportive, there was no increase in attitudes once expectations were exceeded.

Thus far, it has been assumed that Supervisory Style of the Boss and Favorable Self-Perception subscale scores are *causative* variables. Vroom (1964, pp. 112–113), however, has suggested an alternative explanation and expressed the need for joint studies to clarify the relationship. Specifically, he suggested that should job satisfaction be found to be jointly influenced by supervisory style and a personal variable, then more credence could be given to the claim that a supportive style causes changes in job satisfaction, as opposed to the alternative explanation that satisfied subordinates elicit supportive behavior from their supervisor. Since Job Satisfaction subscale scores were found to vary jointly with Supervisory Style of the Boss and Favorable Self-Perception subscale scores, there is evidence in support of the assumption of causation.

REFERENCES

CATTELL, B., & BAGGALEY, R. The salient variable similarity index for factor matching. *The British Journal of Statistical Psychology*, 1960, **13** (1), 33–46.

CLYDE, D. J., CRAMER, E. M., & SHERRIN, R. J. Multivariate analysis of variance. Unpublished manuscript, University of Iowa Computer Center, 1969.

GILMER, R. S. A factorial approach to the life history correlates of intellectual change. Unpublished doctoral dissertation, Purdue University, 1963.

LIKERT, R. *New patterns of management*. New York: McGraw-Hill, 1961.

LIKERT, R. *The human organization: Its management and value*. New York: McGraw-Hill, 1967.

MORRISON, R. F. Factorial life history antecedents of industrial research performance. Unpublished doctoral dissertation, Purdue University, 1962.

MYERS, M. S. Conditions for manager motivation. *Harvard Business Review*, 1966, **44** (1), 58–71.

THOMPSON, D. E. Job satisfaction in administration: An investigation of the joint effects of supervisory style and subordinate self-perception. Unpublished doctoral dissertation, University of Iowa, 1969.

THOMSON, R. W. A factor analysis of the life history correlates of engineering interests. Unpublished doctoral dissertation, Purdue University, 1964.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

(Received May 18, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 353–358

INFLUENCE OF RACE ON THE MANIFESTATION OF LEADERSHIP[1]

JAMES R. FENELON[2]
University of Texas

AND

EDWIN I. MEGARGEE[3]
Florida State University

The question of whether racial differences influence the manifestation of dominance (Do) was investigated by pairing black and white women who were high and low on the California Psychological Inventory Do scale and having them interact in a task in which one had to lead and the other follow. Groups were constituted of the following pairs of college women: (*a*) high- and low-Do whites—Group 1; (*b*) high- and low-Do blacks—Group 2; (*c*) high-Do white and low-Do blacks—Group 3; and (*d*) high-Do blacks and low-Do whites—Group 4. The rate of leadership assumption by the high-Do white women paired with the low-Do black women in Group 3 was significantly lower than the rate of leadership assumption by the high-Do *S*s in the other groups. Analysis of the decision-making process suggested that this stemmed from the reluctance of the high-Do white girls to assume leadership over the low-Do black girls coupled with increased assertiveness on the part of the low-Do black girls when paired with a white partner.

In applied settings, the usefulness of a psychological test is often judged by how well it can predict future behavior (Meehl, 1959). However, most of the research literature on such tests is based on studies of their concurrent rather than their predictive validity (Myers, 1950). For example, the California Psychological Inventory (CPI) Dominance (Do) scale is used in secondary schools, colleges, and industry to forecast who is most likely to rise to a position of leadership and authority (Gough, 1968, 1969a, 1969b). However, most of the evidence for its validity rests on concurrent validity studies that have shown that individuals who are already in a position of leadership—student leaders, military officers, business executives, and the like—obtain significantly higher scores than nonleaders (Bogard, 1960; Carson & Parker, 1966; Gough, 1969a, 1969b; Johnson & Frandsen, 1962; Knapp, 1960; Rawls & Rawls, 1968).

A series of studies on the predictive validity of the Do scale has illustrated one of the basic problems in predicting behavior from psychological test scores: while behavior is a function of both personality and situational factors, tests assess only the personality variables. In the first study, Megargee, Bogart, and Anderson (1966) paired men with high-Do scores with men with low-Do scores and had them work together on a simulated industrial task. Although no specific mention of leadership or role differentiation was made in the instructions, it seemed clear to the investigators that the task required one person to lead and the other to follow. Despite this, only 14 of the 25 high-Do *S*s (56%) assumed the leader role, a difference that was not significant. The study was repeated with the instructions changed so as to stress leadership and emphasize the importance of the leader's role. Given this change, 18 of the 20 high-Do *S*s tested (90%) assumed the leader role ($p < .001$).

If the instructions could make such a difference, it seemed likely that other factors could also influence whether or not high-Do people assumed leadership in various situations. A program of research was undertaken to identify the situational factors that facilitated and inhibited the assumption of leadership by people high in dominance as measured by the CPI Do scale.

The first of these studies (Megargee, 1969) investigated the influence of traditional sex roles on the assumption of leadership by high-Do *S*s. Four groups were formed: (*a*) high-Do men paired with low-Do men, Group 1; (*b*) high-Do women paired with low-Do women, Group 2; (*c*) high-Do men paired with low-Do women, Group 3; and (*d*) high-Do women paired with low-Do men, Group 4. It was expected that the high-Do partner would usually assume the leader role in Groups 1, 2, and 3, where cultural sex role prescriptions did not interfere with the manifestation of dominance. In Group 4, however, it was predicted that the high-Do women who were paired with low-Do men would assume leasership significantly less often because of the tradition of masculine dominance in American society. This proved to be the case. In two studies using two quite different tasks, no more than 25% of the high-Do women paired with low-Do men assumed leadership while in the other groups the high-Do partner assumed leadership 62%–90% of the time.

---

[1] This article is based on a master's thesis submitted to the Psychology Department, University of Texas at Austin in August 1966 by J. R. Fenelon under the direction of E. I. Megargee.

[2] Now at David Grant Hospital, Travis Air Force Base, California.

[3] Requests for reprints should be sent to Edwin I. Megargee, Department of Psychology, Florida State University, Tallahassee, Florida 32306.

The present study further explores how cultural role expectations may influence the prediction of leadership using the Do scale. Whereas Megargee (1969) studied what happens when high- and low-Do men and women are paired and asked to assume leader or follower roles, this study applies the same paradigm to two other groups that have traditionally occupied dominant and submissive roles relative to each other in American society: whites and blacks. It was expected that in the homogeneous racial groups (i.e., high-Do white–low-Do white and high-Do black–low-Do black) the high-Do partners would usually assume the leader role. The question was what would happen in the biracial groups. Would the dominance of white over black common in the South result in a lower rate of leadership assumption by the high-Do blacks? There was reason to believe that it might. Megargee's (1969) study had shown that conflict with stereotyped cultural roles inhibited the expression of dominance by high-Do white women. Prior research on the interaction of blacks and whites had shown that Negro *S*s were inhibited and complained and made more errors when working with whites (Burstein & McRae, 1962; Katz & Benjamin, 1960; Katz & Cohen, 1962; Katz, Goldston, & Benjamin, 1958; Katz & Greenbaum, 1963; Katz, Roberts, & Robinson, 1965). While these studies did not deal with emergent leadership per se, they did suggest a pattern of passivity that would lead one to expect that high-Do black *S*s would be less likely to lead low-Do white partners. Similarly, it seemed likely that a low-Do white *S* might behave more assertively when paired with a high-Do Negro than when paired with a white partner because following a Negro leader might well be threatening to southern white *S*s.

On the other hand, no particular conflict between personality and cultural role expectation was anticipated in the high-Do white–low-Do black situation; hence it was anticipated that the high-Do white *S*s would usually lead the low-Do black *S*s.

METHOD

Subjects. The *S* selection was a major problem in this investigation because of the relatively small number of Negro students enrolled at the University of Texas at Austin at that time. This meant that the primary source of *S*s, the introductory psychology classes, could not be used; instead it was necessary to recruit *S*s in the integrated cooperative housing units at the University. The effect of this was threefold: (*a*) it dictated the use of women rather than men as *S*s because most of the blacks on campus were female, and since the men did not have to live in University-approved quarters as did the women; (*b*) it meant that the sample of whites consisted of those willing to live in integrated housing; (*c*) as a result of the recruiting and the statistical improbability of a white student "just happening" to be paired with a black student on the basis of chance, it is likely that most *S*s surmised that racial attitudes as well as leadership were under investigation.

A 113-item test labeled the Gough Inventory which consisted of all the items from the CPI Do, Cm (Communality), and Gi (Good Impression) scales was administered to 72 black and 119 white women students at the University of Texas at Austin. The overall median Do scale *T* score was 53; *S*s at or above the median were classified as high Do; those below as low Do. Pairs were then formed; in each pair the high Do *S*'s *T* score was at least 20 points higher than that of his low-Do partner. Four groups were formed with 15 pairs of *S*s in each: Group 1 consisted of high-Do whites and low-Do whites, Group 2 of high-Do blacks and low-Do blacks, Group 3 of high-Do whites and low-Do blacks, and Group 4 of high-Do blacks and low-Do whites. Each *S* was contacted by phone, offered $1.00 for participating, and given an appointment to report to the laboratory.

Task and apparatus. The task and apparatus were identical to those used by Megargee (1969). The task consisted of a simulated clerical task with well-defined leader and follower roles in which the importance of the leader's role was stressed. Megargee found that this task, which seemed more appropriate for use with female *S*s, gave results identical to the more masculine simulated industrial task devised by Megargee, Bogart, and Anderson (1966).

The setting consisted of a table with a screen in the middle and chairs on either side. The leader's chair was an executive-type swivel chair with arms. In front of it rested a leather-covered loose-leaf binder containing the material to be dictated. A sign on the screen above the binder read "leader's side." The follower's chair was a straight-backed wooden chair. She had no binder for her papers and on her side of the screen a sign read "follower's side."

The leader's notebook contained a page from the Stroop Color-Word Test (Stroop, 1935) which had the names of four colors, each printed in ink of a different color such as the word "red" printed in blue ink. The follower was supplied with a mimeographed form on which the initial letters of the four color names were printed, once for each word on the leader's sheet.

Procedure. Each pair of *S*s was led to the testing room where the following instructions were read with the italicized words emphasized:

> This is a study of the relation between the Gough Inventory and *leadership* under stress. One important aspect of *leadership* is the ability to concentrate, remain calm, and accurately give directions to a subordinate. This is what is required by this task. The *leader* will have to absorb information and rapidly pass it on to the *follower*.
>
> The information to be transmitted is the color of the ink on a series of words. This is the *leader's* side of the table. On the table is a sheet of paper on which are printed the names of four colors: blue, green, orange, and red, as you can see in the *leader's* sample. Four different colors of ink—blue, green, orange, and red— have been used in printing the color names. As you can see, the word "orange" is printed in blue ink. The job of the *leader* is to tell the *follower* the color ink used, ignoring the printed word. Since the first word is printed in blue ink, the *leader* should call "blue"; since the second word is printed in orange ink, she should call out "orange," and so on.
>
> (The *E* moves around table.)
>
> This is the *follower's* side of the table. The *follower* must record the information given her by the *leader*. The *follower's* sheet of paper has the letters B, G, O, and R printed, once for each word on the *leader's* sheet. As the *leader* calls out the colors of the ink, the *follower* must record the information by crossing out the letter corresponding to the color, B for blue, G for green, O for orange, and R for red.
>
> This sample has been filled out to correspond to the *leader's* sample. Remember the first word on the *leader's* sample was "orange" printed in blue ink; therefore the *leader* should have said "blue" and the *follower* should cross out the letter B. (If not clear give more examples.)
>
> Thus the *leader* must transmit as much information to her subordinate as possible in the time allowed, as if giving orders in a crisis. The *follower* must record what her *leader* tells her.
>
> It is up to you to decide who will be the *leader* and who will be the *follower*. When you have decided, take your seats at the *leader* or *follower's* sides of the table. When the *leader* is sure the *follower* is ready to record the information, she should call out "start," turn over the sample page, and start calling out the names of the different-colored inks. When the *leader* calls out "start," I shall begin timing you; the *leader* will have 90 sec. to transmit as much information as possible.
>
> Any questions?
>
> OK. Decide who will be *leader* and who will be *follower* and take your positions.

Of course, the investigators were not interested in the accuracy or speed of dictation but simply in which *S* ended up in the leader's chair and the process whereby this decision was reached. The purpose of the task was simply to provide a setting in which the high-Do *S* could manifest her dominance by assuming the leader role if she so chose.

After reading the instructions, *E* turned on a tape recorder which recorded the discussions that preceded the choice of leader. Later *E* transcribed these records, noting any additional nonverbal behavior such as an *S* simply sitting down in one position or shrugging when asked her opinion.

RESULTS

Decision time. It had been expected that intra- and possibly interpersonal conflict would be engendered by the biracial situation. Because of this it was predicted that the amount of time taken for the partners to arrive at a decision would be significantly longer in the biracial than in the uniracial groups. This proved to be the case; the mean decision-making time in the biracial groups was 9.8 sec. which proved to be significantly longer than the uniracial groups' mean of 5.8 sec. ($F = 4.65$, $df = 1, 56$, $p < .05$).

Assumption of leader role. The number of high-Do *S*s assuming the leader and follower roles in each of the four groups is presented in Table 1. In both of the two uniracial groups, 80% of the high-Do *S*s assumed the leader role. This was consistent with the expectation that a majority of the high-Do *S*s in the uniracial groups would assume leadership, and it replicated Megargee's (1969) finding that 75% of his high-Do white women *S*s assumed the leader role when paired with low-Do white women *S*s.

In the biracial groups, it had been expected that most of the high-Do white women in Group 3 would become leaders. This did not prove to be the case since only 33% of these women ended up in the leader's chair. In Group 4, a low rate of leadership assumption had been expected on the part of the high-Do black women who were paired with low-Do whites. This, too, proved to be in error; 73% of the high-Do black women assumed the

TABLE 1

umber of High-Do *S*s Assuming Leader and Follower Roles in Each Group

Role Adopted	Group 1 (High white-Low white)	Group 2 (High black-Low black)	Group 3 (High white-Low black)	Group 4 (High black-Low white)
Leader				
N	12	12	5	11
%	80	80	33	73
Follower				
N	3	3	10	4
%	20	20	67	27
N	15	15	15	15

Note.—$\chi^2 = 10.2$, $p < .02$.

leader role. These differences in the rate of leadership assumption in the four groups were statistically significant ($\chi^2 = 10.2$; $p < .02$).

Decision-making process. Finally the records of the verbal and nonverbal interactions during the decision-making process were analyzed. While the cell entries were generally too small to allow for meaningful statistical tests, it was hoped that this analysis would provide fruitful hypotheses to help explain the results. The methods of analysis followed those described by Megargee (1969). It was first noted which *S* made the final decision; this information is recorded in the first row of Table 2. Next, the behavior of the partner who made the decision was studied to determine whether he appointed himself or his partner leader. These data are presented in the second and third rows of Table 2. The behavior of those who allowed their partner to decide was also examined and a record made of those who actively said in effect, "You decide" and those who simply said nothing. The proportion of high- and low-Do *S*s who said nothing is presented in the fourth row of Table 2.

DISCUSSION

The first objective of the present program of research was to explore how situational factors can interfere with the prediction of behavior on the basis of personality test scores. Megargee, Bogart, and Anderson (1966) demonstrated that the degree to which the instruction emphasizes the relevance of leadership to the task can be of crucial importance in determining whether or not a high-Do *S* assumes the leader role. The next study (Megargee, 1969) showed that while high-Do *S*s will generally lead low-Do *S*s, this is not true when it involves a conflict with the cultural role expectation that males should dominate females.

The present study extended these findings by showing that racial differences between the partners can also significantly interfere with predictions based on test scores alone. Moreover, as has often occurred in these studies, the effect was not that which had been expected. It was not the high-Do black women in Group 4 but the high-Do white women in Group 3 who manifested a low rate of leadership assumption when paired with a partner of a different race.

TABLE 2

Analysis of Decision-Making Process

No. *S*s	Group 1	Group 2	Group 3	Group 4
Making decision				
High Do				
N	7	8	6	6
%	47	53	40	40
Low Do				
N	8	7	9	9
%	53	47	60	60
Appointing self leader				
High Do				
N	5	6	0	4
%	71	75	0	67
Low Do				
N	1	1	4	2
%	12	14	44	22
Appointing partner leader				
High Do				
N	2	2	6	2
%	29	25	100	33
Low Do				
N	7	6	5	7
%	88	86	56	78
Deferring by saying nothing				
High Do				
N	8	6	5	8
%	100	86	56	89
Low Do				
N	7	7	4	3
%	100	88	67	50

What happened in these biracial groups? As in any experiment, there are a number of possible ex post facto explanations. However, close scrutiny of the data in Table 2 suggests that the most plausible explanation is that the high-Do white *S*s in Group 3 were reluctant to assume the leader role vis-à-vis their low-Do Negro partners. Moreover, the low-Do black women in Group 3 tended to be more assertive than the low-Do *S*s in the other groups. Let us examine the empirical basis for these speculations.

The first question that was asked was whether the four groups differed in respect to whether the high- or low-Do partner made the decision. The first row in Table 2 shows that no such differences were found. Overall, the final decision was made by the low-Do partner 55% of the time. The individual groups did not differ significantly from this average.

Turning to the behavior of *S*s who made the decisions, some interesting differences can be found. In this experimental situation, when the high-Do partner makes the decision, he typically appoints himself leader (Megargee, 1969). This was the pattern in both the uniracial groups and in the high-Do black–low-Do white group. However, while $\frac{2}{3}$ or more of the high-Do *S*s in Groups 1, 2, and 4 appointed themselves leaders when they made the decision, none of the six high-Do white decision-makers in Group 3 appointed herself leader. Instead, all six of these girls told their low-Do black partner to lead. The presence of a black partner not only did not increase the assertiveness of the high-Do white girls, but it instead made them more submissive, at least overtly.

This is only part of the story. The other part is the behavior of the low-Do partners. In most cases the low-Do partner made the decision. The typical scenario was for the high-Do partner to remain silent and look quizzically at her low-Do partner. When the decision was thus thrust on her, the low-Do partner typically responded by saying, "You be the leader." Not so in Group 3. In Group 3, 9 of the 15 high-Do white partners deferred the decision to their low-Do black partners. However, instead of remaining silent, a ploy which apparently put some pressure on the low-Do partner to nominate the high-Do partner as leader, 44% of the high-Do white girls who deferred the decision actively encouraged their black partners to decide by saying, "Go ahead, you decide" or some such statement.[4] Given this encouragement 44% of the low-Do black girls in Group 3 appointed themselves leader when making the decision, while only 12% to 22% of the low-Do partners appointed themselves leaders in the other groups. There is no way of determining whether the greater assertiveness of the low-Do black girls in Group 3 was because the high-Do white girls actively encouraged them to make the decision or whether the biracial situation simply elicited more assertive behavior.

The pattern of the data suggests that in Group 3 the assumption of leadership was less important to the high-Do white girls than was the demonstration of egalitarian attitudes. It is questionable how generalizable this finding is. In the first place, the situation was a temporary one. If the high-Do white girls had been faced with the prospect of having to occupy a subordinate position for 6 mo. or a year, one wonders if they would have been as eager to adopt the follower role. Second, it was probably obvious to the girls in the biracial groups that this was an experiment on racial attitudes as well as leadership. This probably made the expression of tolerance by the white girls or "black power" by the black girls at least as salient as the manifestation of dominance. Moreover, the method of recruiting *S*s in the cooperative dormitories ensured that the white girls selected were drawn from the more liberal elements of the university community. By the same token, the fact that there were so few blacks enrolled in the university at that time suggests that it was the unusual black girl, one with greater fortitude and emotional stamina, who was able to matriculate successfully. The investigators had, in a sense, been expecting the high-Do black girls to "play Uncle Tom." However, by coming to the university they had renounced the "Uncle Tom" role so it is not surprising in retrospect that they also asserted themselves in the experimental situation. Instead, it was the white girls who adopted the submissive role is-à-vis their black partners, so that it was "Miss Anne" who played "Uncle Tom."

---

[4] It may also be noted that in Group 4, 50% of the white girls who deferred to their black partners actively encouraged their partners to decide.

Further research is, of course, necessary to test the validity of some of these speculations. One such study has already been carried out by Rubinroit (1970). Rubinroit set out to determine whether the results of the present investigation could be attributed to the fact that the study was conducted in a university setting. Rubinroit went to a lower class high school in which approximately 15% of the students were white, 25% black, and 60% Mexican-American. In this setting there would be no reason for members of biracial groups to suspect that racial attitudes were being investigated. After studying all nine combinations (high-Do whites paired with low-Do whites, low-Do blacks paired with low-Do Mexican-Americans, high-Do blacks paired with low-Do whites, low-Do blacks, etc.) it became apparent that in this thoroughly integrated setting, differences in dominance as assessed by the CPI Do scale were the primary predictors of leadership.

REFERENCES

BOGARD, H. M. Union and management trainees—A comparative study of personality and occupational choice. *Journal of Applied Psychology*, 1960, **44**, 56–63.

BURSTEIN, E., & MCRAE, A. Some effects of shared threat and prejudice in racially mixed groups. *Journal of Abnormal and Social Psychology*, 1962, **64**, 257–263.

CARSON, G. L., & PARKER, C. A. Leadership and profiles on the MMPI and CPI. *Journal of College Student Personnel*, 1966, **7**, 14–18.

GOUGH, H. G. An Interpreter's syllabus for the California Psychological Inventory. In P. McReynolds (Ed.), *Advances in psychological assessment.* Vol. 1. Palo Alto, California: Science and Behavior Books, 1968.

GOUGH, H. G. A leadership index on the California Psychological Inventory. *Journal of Counseling Psychology*, 1969, **16**, 283–289. (a)

GOUGH, H. G. *Manual for the California Psychological Inventory.* (Rev. ed.) Palo Alto, California: Consulting Psychologists Press, 1969. (b)

JOHNSON, R. T., & FRANDSEN, A. N. The California Psychological Inventory profile of student leaders. *Personnel and Guidance Journal*, 1962, **41**, 343–345.

KATZ, I., & BENJAMIN, L. Effects of white authoritarianism in biracial work groups. *Journal of Abnormal and Social Psychology*, 1960, **61**, 448–456.

KATZ, I., & COHEN, M. The effects of training Negroes upon cooperative problem solving in biracial teams. *Journal of Abnormal and Social Psychology*, 1962, **64**, 319–325.

KATZ, I., GOLDSTON, J., & BENJAMIN, L. Behavior and productivity in biracial work groups. *Human Relations*, 1958, **11**, 123–141.

KATZ, I., & GREENBAUM, C. Effects of anxiety, threat, and racial environment on task performance of Negro college students. *Journal of Abnormal and Social Psychology*, 1963, **66**, 562–567.

KATZ, I., ROBERTS, S., & ROBINSON, J. Effects of task difficulty, race of administrator, and instructions on digit symbol performance of Negroes. *Journal of Personality and Social Psychology*, 1965, **2**, 53–59.

KNAPP, R. R. A reevaluation of the validity of MMPI scales of dominance and social responsibility. *Educational and Psychological Measurement*, 1960, **20**, 381–386.

MEEHL, P. E. Some ruminations on the validation of clinical procedures. *Canadian Journal of Psychology*, 1959, **13**, 102–128. (Reprinted in E. I. Megargee [Ed.], *Research in clinical assessment.* New York: Harper & Row, 1966.)

MEGARGEE, E. I. The influence of sex roles on the manifestation of leadership. *Journal of Applied Psychology*, 1969, **53**, 377–382.

MEGARGEE, E. I., BOGART, P., & ANDERSON, B. J. The prediction of leadership in a simulated industrial task. *Journal of Applied Psychology*, 1966, **50**, 292–295.

MYERS, C. R. Prediction in clinical psychology. *Canadian Journal of Psychology*, 1950, **4**, 97–108. (Reprinted in E. I. Megargee [Ed.], *Research in clinical assessment.* New York: Harper & Row, 1966.)

RAWLS, D. J., & RAWLS, J. R. Personality characteristics and personal history data of successful and less successful executives. *Psychological Reports*, 1968, **23** (3, Pt. 2), 1032–1034.

RUBINROIT, C. Leadership in dyadic groups as a function of dominance and ethnic combination. Unpublished doctoral dissertation, University of Texas, 1970.

STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643–662.

(Received April 30, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 359–363

INFORMATION SEEKING IN A NATURAL STRESS SITUATION[1]

DAVID T. A. VERNON[2]

University of Colorado

Hospitalized tuberculosis patients under the care of informative and uninformative physicians were compared with respect to their use of books in the patient library. The two groups did not differ in the proportion of general book readers. However, in line with the findings of prior laboratory experiments on information choice and information seeking in stress situations the patients of uninformative physicians tended to seek out books about tuberculosis and its treatment more frequently than did the patients of informative physicians.

The findings of several recent studies suggest: (*a*) that persons in potentially stressful situations in which injury or pain are threatened generally choose to have information about the nature and timing of a stress stimulus when a choice is offered to them, (*b*) that they actively seek such information if it is not offered and, (*c*) that they profit from obtaining it. These tendencies appear to be true even when the information does not seem to be useful in reducing the magnitude of the stress stimulus in an objective way.

Data on the first of these issues are available from the laboratory experiments of Lockhard (1963); Perkins, Levis, and Seymann (1963); and Lanzetta and Driscoll (1966). Lanzetta and Driscoll found that persons prefer to have information about the incidence of an unavoidable shock when given a choice. Similarly, both Lockhard and Perkins et al. found that rats preferred that cage compartment in which a signal light preceded the advent of shock over the compartment without a signal light or the compartment in which the light followed the shock.

With respect to the closely related second issue—that persons in potentially stressful situations actively seek information when it is not offered—less data are available. The field observations of Roth (1963) describe how tuberculosis patients on occasion go to considerable lengths to get information about their condition and their treatment. Pervin (1963) and, to a somewhat lesser degree, Haggard (1943) found that *S*s in laboratory experiments which involved unavoidable electric shock actively tried to determine the signals or pattern of signals (lights or tachistoscopically presented words) which cued the shocks.

Experimental data on the third issue—on the utility of such information for modifying the psychological impact of the stress stimuli, not their objective magnitude—appear to be reasonably consistent. In research in a closely related area, Carlsmith and Aronson (1963) found that a disagreeably bitter solution was rated as tasting more bitter when it was not expected. Pervin's (1963) data indicated a similar phenomenon—that is, persons rated a predictable electric shock as being preferable and less anxiety provoking than an unpredictable shock. Haggard (1943) found that *S*s in his experiment who developed "cognitive clarity" (who were able to identify the specific words which cued an electric shock) were able to tolerate an electric shock better than persons for whom the situation remained ambiguous. The former showed generally lower galvanic skin responses, more accurate estimates of session length, and less disintegration of verbal-motor coordination. The data of Lanzetta and Driscoll (1966), while agreeing with the general trend of the above studies, must be considered less conclusive since supporting interpretations could be reached only from internal analyses and not from the main treatment group comparisons.

---

[1] The author wishes to thank Veterans' Administration Hospital, Hines, Illinois for its cooperation and support during the conduct of the research. The present article is an elaboration of a segment (Vernon, 1967, Appendix G, pp. 122–123) of research conducted in partial fulfillment of requirements for the PhD degree at the University of Chicago. It was carried out under the supervision of F. L. Strodtbeck, University of Chicago.

[2] Requests for reprints should be sent to David T. A. Vernon, Department of Psychology, University of Colorado, Boulder, Colorado 80302.

The field research of Janis (1958) with
rgery patients, although purely correlational
nature, also tends to support the proposition
at information of the above sort is generally
eneficial, as do numerous problem-oriented
eld experiments on the effectiveness of pre-
aring children and adults for medical pro-
edures (e.g., Dumas, Anderson, & Leonard,
965; Egbert, Battit, Welch, & Bartlett, 1964;
nd the research reviewed by Vernon, Foley,
ipowicz, & Schulman, 1965). However, char-
cteristically these latter studies have been
lesigned in such a way that the effects of
nformation could not be separated from the
effects of other aspects of the experimental
preparation routines.

In addition to reducing the impact of potentially stressful stimuli (e.g., the perceived painfulness of an electric shock or medical procedure as described above), information may also modify the degree of response to the threat of such stimuli. This is evident in the studies of Lazarus and Alfert (1964) and Folkins, Lawson, Opton, and Lazarus (1968), both of which provide data suggesting that responses to a threatening movie may be modified by preparatory information. In the second study this interpretation is based on the fact that the two most effective treatments involved a rather complete description of the film in addition to procedures designed to encourage particular components of desensitization therapy. The work of Nomikos, Opton, Averill, and Lazarus (1968) ostensibly provides evidence that information leads to increased rather than decreased fear. However, as far as the present issue is concerned, these findings must be considered very carefully, as the authors recognized. The average length of time between information and the period of maximum threat was only 12 sec. in this study.

It was the purpose of the present work to investigate the second of the above notions—that is, that pertaining to information seeking—in a natural field setting, the tuberculosis service of a large veterans' administration general hospital.

In the light of the above findings it seemed reasonable that tuberculosis patients under the care of relatively uninformative physicians would be more inclined to actively seek information about the disease and treatment from secondary sources than would patients under the care of informative physicians.

The appropriateness of this expectation for the present setting was further suggested by numerous descriptive emphases on the importance of information to the emotional well-being of tuberculosis patients and by surveys which have demonstrated the phenomenological importance of information to persons in hospitals (Skipper, 1965; Stewart, 1956). The findings of these two studies also suggest that patients (quite reasonably) regard their doctors as the preferred source of information about their disease and treatment.

METHOD

High and Low Information Wards

On the basis of statements about staff doctors made by five patients, four registered nurses, and a clinical psychologist who worked on the tuberculosis service, six of the seven wards of the tuberculosis service were classified as either high information wards or low information wards. At the time of the study, one ward, in serving as a discharge ward, fulfilled a special function for the tuberculosis service and was excluded from the study. The staff doctors in charge of the three high information wards were generally characterized as being relatively free in giving information about tuberculosis and treatment to their patients. They were described as being attentive to their patients' questions, willing to stay after hours to explain test results and treatment procedures to their patients, etc. In contrast, the staff doctors in charge of the three low information wards were described as being relatively uninformative, resentful of patients' requests for information, difficult for patients to contact.

The amount of agreement among the 10 respondents was quite high. The classification of the six physicians was identical in seven cases; each of the remaining three (two patients and a nurse) agreed with the consensus picture in 4 out of 6 of their judgments.

The validity of the classification was also supported by data from a sample of 100 new patients who were tested on seven questions about their own particular cases 1 mo. after admission to the tuberculosis service (Vernon, 1967). This information (e.g., whether the patient's disease was unilateral or bilateral) was of the type that could not be successfully derived from secondary sources of information; it could be judged right or wrong only from each patient's medical records. On the average, the 44 patients on the high information wards had significantly more information (total correct wrong) than did the 56 patients on the low information wards (2.22 vs. 1.09, $t = 3.25$, $.01 > p > .001$).

TABLE 1

DATA ON THE READERSHIP OF TUBERCULOSIS BOOKS AND GENERAL BOOKS IN HIGH AND LOW INFORMATION WARDS

Books	Low information wards				High information wards			
	A	B	C	Total	D	E	F	Total
Tuberculosis								
No. readers	22	21	17	60	9	3	12	24
No. potential readers	242	205	254	701	169	181	157	507
Proportion	.09	.10	.07	.09	.05	.02	.08	.05
General								
No. readers	18	12	9	39	14	6	16	36
No. potential readers	242	205	254	701	169	181	157	507
Proportion	.07	.06	.04	.06	.08	.03	.10	.07

Use of Secondary Sources of Information

Thirteen books in the tuberculosis service library were designated as potential sources of information for tuberculosis patients. These included all autobiographical accounts by former patients and all books written specifically for tuberculosis patients by experts. In addition to the 13 tuberculosis books, 30 general books (both fiction and nonfiction) were selected randomly. From information which was routinely recorded on the circulation cards it was possible to determine the number of tuberculosis book readers and general book readers on each of the six wards over a preceding 2-yr. period during which the medical supervision and organization of the wards was stable.

Comparison of the 1-yr. totals for the six wards indicated that the 2-yr. totals were relatively reliable measures. The correlation between the two 1-yr. totals, corrected by the Spearman-Brown formula, was .86 for the tuberculosis books and .87 for the general books. The measure also possessed the virtue of being nonreactive (Campbell, 1957).

It was assumed throughout that the proportion of patients who were curious about, or were capable of becoming curious about, tuberculosis and treatment was equivalent from ward to ward at the time of their admission to the hospital. This assumption seemed plausible in light of the fact that the patients were assigned to wards on the basis of available bed space, a method which seemed unlikely to bias comparisons between different types of wards.

RESULTS

Table 1 summarizes the essential data of the study—the number of tuberculosis book readers and the number of general book readers on each of the six wards for the 2-yr. period, along with the corresponding proportions. Data on potential readers were abstracted from routine hospital census reports.

The proportions of tuberculosis book readers listed in Table 1 are all quite low, the highest being .10. Relatively few patients read books about tuberculosis and treatment.

The extent to which the patients in the two types of wards differed in the degree to which they read books about tuberculosis was tested by comparing high and low information wards on: (*a*) the mean proportion of tuberculosis book readers, .05 vs .09, and (*b*) the mean proportion of general book readers, .07 vs .06. Statistical tests indicated that the two groups differed significantly with respect to tuberculosis book readers ($Z = 2.11$, $p = .034$, two-tailed test) but not general book readers ($Z = -.05$, $p = .96$, two-tailed test).[3]

DISCUSSION

The findings cited above are in general agreement with the findings of closely related laboratory experiments and anecdotal studies on information seeking (Haggard, 1943; Pervin, 1963; Roth, 1963) and with the implications of laboratory experiments on information choice in potentially stressful situations (Lanzetta & Driscoll, 1966; Lockhard, 1963; Perkins et al., 1963).

These studies, in conjunction with the studies of whether or not information modifies the impact of a potentially stressful stimulus

---

[3] The tests used here were those appropriate to clustered samples (Scott & Wertheimer, 1962, p. 231). Alternative ways of analyzing the data—for example, considering the six wards to be the *S*s and the difference between the number of tuberculosis readers and the number of general readers to be scores—lead to differences at least as significant by statistical criteria as those described here.

(Carlsmith & Aronson, 1963; Haggard, 1943; Janis, 1958; Pervin, 1963; and others), emphasize the importance of information in stress situations and create a consistent, intelligible picture.

Information about the nature and timing of a potentially stressful stimulus is important to most persons who are in a stress situation because, in some undefined way, it increases their ability to withstand its impact. For this reason persons generally will choose to have such information if it is readily available and will actively seek it if it is not. A natural stress situation such as a hospital is complex in that it contains a variety of potential sources of information which are probably differentially desirable to patients. If information is readily available from preferred sources (e.g., physicians) the patients will use it; if it is not, they will turn to less preferred sources (e.g., books about the disease and treatment, and perhaps nurses, aides, other patients, etc.).

It should be emphasized that the present findings are from a natural, field experiment and as such are subject to alternative interpretations which place less emphasis on the role of information per se. For example, it is possible that other aspects of the physician's personalities or treatment philosophies may have been crucial here. Differences among physicians in strictness, in sympathy, or in friendliness toward patients may have influenced the patients' general anxiety and morale. This, in turn, may have fostered differences in the readership of tuberculosis books which were independent of the effects of physician informativeness. While such interpretations cannot be eliminated in the present study, they seem less compatible with the results of prior laboratory studies than that offered above which emphasizes information.

The fact that the use of library books about tuberculosis was not high in general (only about 7% of all patients were tuberculosis book readers) was not surprising in retrospect and was not taken as an indication of the lack of importance of information to these patients in general. The absolute level of this statistic was assumed to be influenced by the patients' use of other sources, both preferred and less preferred or secondary. In addition, several facts about the setting suggested that books might be a relatively unpopular secondary source of information. The preponderance of patients entering the tuberculosis service were of low socioeconomic status. Many had little formal education, were probably poor readers, and perhaps avoided books which they felt were difficult to understand. In addition, many were elderly and had poor vision with or without glasses. Such factors would certainly reduce the overall extent to which such books were used. With respect to the general books, the low proportions presented are probably artifacts. The sample of 30 general books selected represented only .4% of the total number of general books in the library. A larger sample of general books would have increased this proportion considerably.

Finally, we should emphasize what appear to be the rather clear methodological implications of this study for field experiments on the effect of information, psychological preparation, etc. on human responses to stress. These studies have been numerous in the past, particularly in the area of human responses to medical procedures. In the future, investigators attempting to give some persons in a natural stress situation (e.g., some hospital patients) special information about an impending stress should take pains either to control the use of other sources of information, or, alternatively, to assess the effects of these outside sources on the experimental and control groups. The data presented here suggest that the effects of these outside sources cannot be assumed to be equivalent among such groups.

REFERENCES

CAMPBELL, D. T. Factors relevant to the validity of experiment in social settings. *Psychological Bulletin*, 1957, 54, 297–312.

CARLSMITH, J. M., & ARONSON, E. Some hedonic consequences of the confirmation and disconfirmation of expectancies. *Journal of Abnormal and Social Psychology*, 1963, 66, 151–156.

DUMAS, R. G., ANDERSON, B. J., & LEONARD, R. C. The importance of the expressive function in preoperative preparation. In J. Skipper & R. C. Leonard (Eds.), *Social interaction and patient care*. Philadelphia: Lippencott, 1965.

EGBERT, L. D., BATTIT, G. E., WELCH, C. E., & BARTLETT, M. K. Reduction of postoperative pai by encouragement and instruction of patients: study of doctor-patient rapport. *New England Journ of Medicine*, 1964, 270, 825–827.

FOLKINS, C. H., LAWSON, K. D., OPTON, M. E., & LAZARUS, R. S. Desensitization and the experimental reduction of threat. *Journal of Abnormal Psychology*, 1968, **73**, 100–113.

HAGGARD, E. Some conditions determining adjustment during and readjustment following experimentally induced stress. In S. Tompkins (Ed.), *Contemporary psychology*. Cambridge: Harvard University Press, 1943.

JANIS, I. L. *Psychological stress*. New York: Wiley, 1958.

LANZETTA, J., & DRISCOLL, J. M. Preference for information about an uncertain but unavoidable outcome. *Journal of Personality and Social Psychology*, 1966, **3**, 96–102.

LAZARUS, R. S., & ALFERT, E. The short-circuiting of threat by experimentally altering cognitive appraisal. *Journal of Abnormal and Social Psychology*, 1964, **69**, 195–205.

LOCKHARD, J. A. Choice of warning signal or no warning signal in an unavoidable shock situation. *Journal of Comparative and Physiological Psychology*, 1963, **56**, 526–530.

NOMIKOS, M., OPTON, E., JR., AVERILL, J., & LAZARUS, R. Surprise versus suspense in the production of stress reaction. *Journal of Personality and Social Psychology*, 1968, **8**, 204–208.

PERKINS, C., JR., LEVIS, D., & SEYMANN, R. Preference for signal-shock vs. shock-signal. *Psychological Reports*, 1963, **13**, 735–738.

PERVIN, L. A. The need to predict and control under conditions of threat. *Journal of Personality*, 1963, **31**, 570–587.

ROTH, J. A. Information and the control of treatment in a tuberculosis hospital. In E. Freidson (Ed.), *The hospital in modern society*. Glencoe: The Free Press, 1963.

SCOTT, W. A., & WERTHEIMER, M. *Introduction to psychological research*. New York: Wiley, 1962.

SKIPPER, J. K. Communication to the hospitalized patient. In J. Skipper & R. C. Leonard (Eds.), *Social interaction and patient care*. Philadelphia: Lippencott, 1965.

STEWART, B. Problems faced by staff in meeting the psychological needs of tuberculosis patients. In P. Sparer (Ed.), *Personality, stress, and tuberculosis*. New York: International Universities Press, 1956, 300–316.

VERNON, D. T. A., FOLEY, J. M., SIPOWICZ, R., & SCHULMAN, J. L. *The psychological responses of children to hospitalization and illness*. Springfield: Charles C Thomas, 1965.

VERNON, D. T. A. The roles of anticipatory fear and information in tuberculosis patients' responses to hospitalization and illness. Unpublished doctoral dissertation, University of Chicago, 1967.

(Received February 18, 1970)

ournal of Applied Psychology
971, Vol. 55, No. 4, 364–371

DENTIFICATION OF MODERATOR VARIABLES BY DISCRIMINANT ANALYSIS IN A MULTIPREDICTABLE GROUP VALIDATION MODEL[1]

SHELDON ZEDECK[2]

Bowling Green State University

A multivariate discriminant analysis procedure was used to identify moderator variables in a multipredictable group validation (MPGV) model. Results from MPGV also were compared to those of the univariate simple algebraic and absolute difference techniques. Seventeen variables, determined for 418 undergraduate students ($N = 209$ in the experimental group; $N = 209$ in the cross-validation group), were investigated as potential moderators of the relationship between the predictor and the criterion (grade point average). The three techniques revealed moderators different from each other. In addition, number and types of moderators varied according to the composition of the prediction groups in MPGV. Results are discussed with respect to problems with moderator variable techniques.

In an effort to increase predictive efficiency, investigators have modified the classic prediction model. In particular, Dunnette's (1963) model gives considerable attention to interactions which may occur between several factors—predictors, individuals, behaviors in the situation, and consequences of these behaviors relative to the goals of the group. In Dunnette's model, it is necessary to identify relatively homogeneous subsets of predictors, individuals, behaviors, and situations in an examination of the interactions between predictors and criteria.

The basic approach to the identification of homogeneous sets of individuals has involved the identification and use of moderator variables. Banas (1964) defined a moderator variable as a quantitative or qualitative variable which improves the usefulness of a predictor by isolating subgroups of individuals for whom a predictor or set of regression weights is especially appropriate. The original concept of a moderator variable referred to a multivariate curvilinear regression equation using cross products (moderator and predictor) in which the beta weights, instead of being constant, are linear functions of the moderator variables (Saunders, 1955, 1956). The basic notion of moderator variables also is implied in subgrouping variables (Frederiksen & Melville, 1954), population control variables (Gaylord & Carroll, 1948), predictability variables (Ghiselli, 1956, 1960b), referrent variables (Toops, 1959), and modifier variables (Grooms & Endler, 1960).

A common procedure used to identify moderator variables is Ghiselli's (1956, 1960b) technique which involves deviations from the regression line. Correlates of absolute difference scores (/*D*/), between standardized criterion scores (Z_c) and standardized predictor scores (Z_p), are used to identify relatively predictable (low /*D*/ scores) from unpredictable individuals.

Another technique, quadrant analysis (Hobert & Dunnette, 1967), revealed items which distinguished between overpredicted (individuals whose Z_p scores are below their Z_c scores) and predicted (Z_p and Z_c scores are approximately equal), and between underpredicted (individuals whose Z_p scores are below their Z_c scores) and predicted individuals. Identifying differences among unpredictable groups provides enhancement of the understanding of the psychological composition of groups. However, recent comments by Abrahams (1969) and McNemar (1969), indicate limitations in the quadrant analysis method as used by Hobert and Dunnette (1967).

One purpose of this article is to examine the results obtained from a combination of the two aforementioned approaches. A multipredictable group validation (MPGV) involves deviations from the regression line, algebraic differences, and identification of predicted, underpredicted, and overpredicted individuals. Prediction groups are formed by taking the algebraic difference ($\pm D$) between an individual's Z_c score and Z_p score, $\pm D = Z_c - Z_p$. The underpredicted group is composed of individuals with D greater than some a priori value, for example, $> +1$, the overpredicted group with $D < -1$, and the predicted group with $-1 \leq D \leq +1$. Size of the predicted region and number of levels of underpredicted and overpredicted groups could vary depending on the situation.

---

[1] The author wishes to express his thanks to Patricia Cain Smith, C. J. Cranny, Carol A. Vale, and Michael Doherty for their guidance throughout this research.

[2] Requests for reprints should be sent to the author who is now at the Department of Psychology, University of California, Berkeley, California 94720.

A second purpose of this research is to examine a technique which would identify moderator variables. The method of trying out specific hypotheses is slow and expensive because of the tremendous number of possibilities that exist. For the same reasons, the development of moderator scales through item analyses is no more satisfactory. There are no statistics or objective methods that mechanically identify moderators. Furthermore, not enough evidence has accumulated to make it possible to state any general principles about the nature of items or traits that function as moderators.

French (1961) attempted to develop a program for the selection of moderator variables from a large number of variables. The program was designed to identify moderators from 42 potential moderator variables without taking time to compute actual correlations. Indexes, based on the joint distribution, in 3 × 3 contingency tables, of predictor and criterion scores which had been trichotomized, were used in place of correlation coefficients. The technique was unsuccessful because of distortion of the indexes.

Since moderator variables discriminate between those close to the line of regression and those farther away (Lykken & Rose, 1963), it is conceivable that the statistical method of discriminant analysis will identify the variables that contribute to the discrimination of overpredicted, predicted, and underpredicted individuals in the MPGV model. Hence, a discriminant analysis procedure is used for the systematic identification of moderator variables in MPGV.

The third purpose of this research is to compare the results of MPGV with those of simple algebraic and absolute difference techniques. To achieve these ends, several variables are examined to determine their moderator effects on prediction of college grade point average.

METHOD

Subjects. The *S*s were 500 undergraduate students fulfilling a research participation requirement of an introductory psychology course. Because of incomplete predictor and criterion information, sample size was reduced to $N = 418$. The *S*s were randomly assigned to the experimental group, Sample A ($N = 209$), and to the cross-validation group, Sample B ($N = 209$).

Criterion. The criterion measure was student's accumulative grade point average (GPA) compiled prior to the winter 1969 quarter.

Predictor. The predictor was the composite score obtained on the *American College Test* (ACT; American College Testing Program, 1965) administered prior to admission to the university.

Potential moderator variables. Six tests, administered to *S*s in one session, were examined to determine their effects on the relationship between ACT and GPA. The tests and subtests are presented below. These tests and subtests provided scores for 17 potential moderator variables.

Analysis of Relationships (A-R) (Ghiselli, 1960a)
Bell Adjustment Inventory (Bell, 1962)
    Emotionality (E)
    Health Adjustment (He)
    Home Adjustment (Ho)
    Hostility-Friendliness (H-F)
    Masculinity-Femininity (M-F)
    Submissiveness (S)
Culture Fair Intelligence Test ("g") (Cattell, 1963)
Mandsley Personality Inventory (Eysenck, 1962)
    Neurotocism (N)
    Extroversion-Introversion (E-I)
Multiple Affect Adjective Check List (MAACL) (Zuckerman & Lubin, 1965)
    Anxiety (A)
    Depression (D)
    Hostility (H)
Survey of Study Habits and Attitudes (Brown & Holtzman, 1965)
    Delay Avoidance (DA)
    Work Method (WM)
    Teacher Approval (TA)
    Education Acceptance (EA)

Procedure

Multipredictable group validation model. All scores (predictor, criterion, and potential moderators) were converted to standardized Z scores. Overpredicted, predicted, and underpredicted groups were formed based on the algebraic difference score, $\pm D = Z_{GPA} - Z_{ACT}$. Sample A was split into overpredicted, predicted, and underpredicted groups in seven different ways, thereby providing seven sample problems for analysis. The seven problems differed in their a priori interval for the prediction group (e.g., ±.2, ±.5, ±1.0, etc.)

TABLE 1

PREDICTION GROUP INTERVALS

Group	Problem						
	1	2	3	4	5	6	7
Overpredicted	$D < -.5$	$D \leq -1.0$	$D \leq -1.5$	$D < -1.4$ $-1.4 \leq D < -.8$ $-.8 \leq D < -.5$ $-.5 \leq D < -.2$	$D < -1.1$ $-1.1 \leq D < -.4$ $-.4 \leq D \leq -.1$	$D < -1.4$ $-1.4 \leq D < -.9$ $-.9 \leq D < -.5$ $-.5 \leq D < -.2$	$D < -.7$ $-.7 \leq D \leq -.4$
Predicted	$-.5 \leq D < +.5$	$-1.0 < D < +1.0$	$-1.5 < D < +1.5$	$-.2 \leq D \leq +.2$ $+.5 \geq D > +.2$ $+.8 \geq D > +.5$	$-.1 < D < +.1$ $+.4 \geq D \geq +.1$ $+1.1 \geq D \leq +.4$	$-.2 < D < +.2$ $+.5 \geq D \geq +.2$ $+.9 \geq D > +.5$	$-.4 < D < +.4$ $+.7 \geq D \geq +.4$
Underpredicted	$D > +.5$	$D \geq +1.0$	$D \geq +1.5$	$+1.4 \geq D > +.8$ $D > +1.4$	$D > +1.1$	$+1.4 \geq D > +.9$ $D > +1.4$	$D > +.7$

Note.—Levels within problems advance vertically.

and/or the number of levels of underpredicted or overpredicted individuals (i.e., for certain problems, the underpredicted and overpredicted groups were further delineated thereby providing several levels of underpredicted or overpredicted individuals). The parameters differentiating the problems are presented in Table 1.

For each problem, a one-way multivariate analysis of variance (MANOVA), with the 17 potential moderators used as the dependent vector random variable, was performed in Sample A. A significant prediction group main effect, tested by Wilks' lambda criterion, indicates that at least one of the discriminant functions (with standardized coefficients) is significant.

To identify potential moderators within each problem, a stepwise application of multivariate analysis of covariance (MANOCOVA) was used. The initial MANOCOVA treated the variable associated with the highest discriminant function coefficient as the covariate and the remaining 16 as dependent. A significant main effect indicated that some discriminating power still remained among the 16 dependent variables. Each succeeding MANOCOVA added the variable with the next highest weight to the set of covariates and removed it from the dependent set. The procedure was repeated until the dependent set no longer produced a significant main effect.

Once the moderators were identified, a new discriminant function using only those moderators was found and a discriminant score for each S was computed; $DS_i = \sum w_j x_{ij}$, where the discriminant score for the individual i is DS_i, w_j is the coefficient of the j^{th} variable, and x_{ij} is the individual i's score on the j^{th} variable.

Classification of individuals in Sample B was established in two ways. First, cutoff discriminant scores were obtained for each a priori prediction group in Sample A based on England's (1961) Index of Differentiation which minimizes the number of misclassifications. Thus, in Classification Method I, an S in Sample B was placed in a prediction group based on cutoff scores established in Sample A. Tetrachoric correlation coefficients were computed between observed and predicted group classification of Ss. For computational purposes, prediction groups were dichotomized as indicated in the Results section.

The second classification method was based on a chi-square technique suggested by Cooley and Lohnes (1962). The procedure, applied to Ss in Sample B, places an individual into a group depending on the deviation of his discriminant score from the appropriate group mean discriminant score determined in Sample A. Again, tetrachoric correlation coefficients were computed to assess the effectiveness of the moderator in discriminating multipredictable groups in the cross-validation sample.

To assess the effectiveness of the moderators on validity, validity coefficients of the predictor, ACT, were computed for each prediction group in Sample B for both classification techniques. It was hypothesized that validity coefficients of ACT would be higher for the prediction groups in Sample B than the overall coefficient in Sample A.

Absolute difference model. Each potential moderator was correlated with the absolute difference $(/D/)$

between the criterion, GPA, and the predictor, ACT. If a significant correlation was obtained in Sample A, then the sample was split into $\frac{1}{3}$ (and $\frac{2}{3}$) high scorers and $\frac{2}{3}$ (and $\frac{1}{3}$) low scorers on the potential moderator variable, and validity coefficients were obtained for each fractionated group. Significantly different coefficients indicated the existence of a moderator. Results were cross-validated in Sample B.

Algebraic difference model. Each potential moderator was correlated with the algebraic difference ($\pm D$) between the criterion, GPA, and the predictor, ACT. If a significant correlation was obtained in Sample A, then the sample was trichotomized. Differential validity coefficients of the predictor for the subgroups indicated the existence of a moderator variable. Results were cross-validated in Sample B.

RESULTS[3]

The validity coefficient of ACT as a predictor of GPA was $.52 (p < .01)$ and $.51 (p < .01)$ for Sample A ($N = 209$) and Sample B ($N = 209$), respectively.

Multipredictable Group Validation Model

Assessment of potential moderator variables. Seven one-way MANOVAs were computed for the problems outlined in Table 1. The main effect was prediction group, and the levels within the factor varied according to the problem. Scores for the 17 potential moderator variables constituted the dependent vector variable in the analyses. A significant ($p < .01$ in each case) main effect was found in Problems 1–6; Problem 7 did not yield a significant main effect and thus was excluded from further examination.

For each problem, the first discriminant function was analyzed by MANOCOVAs to determine which variables were contributing significantly to the discrimination.

To confirm the significance of the contributing variables, one-way MANOVAs were conducted for each problem with all significant variables treated as the dependent vector variable. Results demonstrated that all main effects were significant ($p < .01$ in each case). The weights of the new discriminant function for each problem are presented in Table 3.

TABLE 2

DISCRIMINANT WEIGHTS FOR THE SEVENTEEN POTENTIAL MODERATOR VARIABLES IN EACH PROBLEM

Variable[a]	Problem					
	1	2	3	4	5	6
E-I	.031	−.076	.243	.106	.606[b]	.363
N	−.445	.090	.398	.131	.340	.290
DA	−.398	.251	−.103	−.082	−.036	.082
WM	−.297	−.145	−.044	−.263	−.034	−.159
TA	.138	−.032	−.164	.219	.293	.346
EA	−.078	.243	.626[b]	.554[b]	.493[b]	.366
"g"	.475[b]	−.334	.016	.138	−.037	.064
Ho	−.158	.116	.084	−.004	−.182	−.014
He	.484[b]	−.374	−.143	−.068	−.055	−.002
S	.138	−.644[b]	−.494	−.294	.162	−.008
E	.291	−.091	−.016	.176	−.227	−.038
H-F	−.465[b]	.297	−.078	.086	.047	−.089
M-F	.058	−.099	−.189	−.425	−.347	−.409[b]
A	−.630[b]	.923[b]	.395	.236	.071	.182
D	.064	−.165	.316	−.221	.418	.131
H	.641[b]	−.486[b]	−.478	−.167	−.146	−.255
A-R	.353	−.380[b]	−.593[b]	−.753[b]	−.664[b]	−.737[b]

[a] See text for explanation of abbreviations.
[b] Significant weights of the discriminant function for the problem as determined by MANOCOVAs.

Effectiveness of moderators in discriminating prediction groups: Classification Method 1. In Sample A, a distribution of discriminant scores was obtained for each level within a problem. Examination of the distributions resulted in two modifications of procedure. First, for problems with more than one level of overpredicted or underpredicted intervals there was considerable overlap between groups. Thus, *S*s within all levels of underpredicted groups were combined into a single underpredicted group and, similarly, all levels of overpredicted groups were combined into a single overpredicted group. Second, for several problems with only three groups, there still existed considerable overlap between groups. Hence, to use England's (1961) Index of Dif-

TABLE 3

DISCRIMINANT WEIGHTS FOR SETS OF CONTRIBUTING VARIABLES IN EACH PROBLEM

Variable[a]	Problem					
	1	2	3	4	5	6
"g"	.636					
He	.580					
H-F	−.725					
A	−.633	−.892				
H	.751	.568				
S		.664				
A-R		.593	1.004	.993	−.870	.701
EA			.599	−.699	.730	
E-I					.491	
M-F						.708

[a] See text for explanation of abbreviations.

---

[3] Tables of intercorrelations and summary of results of MANOVA and MANOCOVA can be obtained from the author. (See address in Footnote 2.)

TABLE 4

CORRELATIONS BETWEEN GPA AND ACT FOR PREDICTION GROUPS IN SAMPLE B BASED ON INDEX OF DIFFERENTIATION

Problem	Prediction group[a]		
	Overpredicted	Predicted	Underpredicted
1	.49 ($N = 173$)		.63 ($N = 36$)
2	.51 ($N = 147$)		.50 ($N = 62$)
3	.49 ($N = 147$)		.53 ($N = 62$)
4	.55 ($N = 44$)	.51 ($N = 107$)	.55 ($N = 58$)
5	.51 ($N = 114$)	.57 ($N = 30$)	.50 ($N = 65$)
6	.53 ($N = 124$)		.49 ($N = 85$)

[a] Values centered in the overpredicted-underpredicted column are common to both groups.

ferentiation, the predicted group was combined with one of the unpredicted groups in Problems 1, 2, 3, and 6.

Based on the cutoff discriminant scores obtained in Sample A, *S*s in Sample B were placed onto prediction groups. There were significant tetrachoric correlations for Problems 3 and 6 ($p < .01$ and $p < .05$, respectively).

Effectiveness of moderators in discriminating prediction groups: Classification Method 2. The chi-square technique (Cooley & Lohnes, 1962) resulted in a second set of discriminant cutoff scores. None of the tetrachoric correlation coefficients revealed a significant relationship.

Effectiveness of moderators on validity: Groups based on Index of Differentiation. Table 4 presents the correlation coefficients between ACT and GPA for the prediction groups within each problem. None of the coefficients for the prediction groups within any problem differed significantly from .52 (overall validity coefficient for Sample A).

Effectiveness of moderators on validity: Groups based on chi-square classification. Table 5 presents the correlation coefficients between ACT and GPA for the prediction groups within each problem. Again, none of the coefficients for the prediction groups within any problem differed significantly from .52.

Absolute Difference Technique

The correlations between $/D/$ and Teacher Approval (TA) ($r = -.17$, $p < .05$) and $/D/$ and Education Acceptance (EA) ($r = -.15$, $p < .05$) were significant in Sample A. For each of the two potential moderator variables (TA and EA), two fractionations were made: (*a*) $\frac{1}{3}$ low and $\frac{2}{3}$ high scorers, and (*b*) $\frac{2}{3}$ low and $\frac{1}{3}$ high scorers. Table 6 presents the correlations between GPA and ACT for each of the fractionated groups. The difference between validity coefficients for the $\frac{1}{3}$ low and $\frac{2}{3}$ high TA scorers and for the $\frac{1}{3}$ low and $\frac{2}{3}$ high EA scorers was significant ($p < .05$ in both cases).

Neither of the differences in correlations between fractionated groups was significant in the cross-validation sample.

Algebraic Difference Technique

The correlations between $\pm D$ and the Culture Fair Intelligence Test ("g") ($r = -.15$, $p < .05$), $\pm D$ and Masculinity-Femininity (M-F) ($r = -.22$, $p < .01$), and $\pm D$ and the Analysis of Relationships (A-R) ($r = -.29$, $p < .01$) were significant. For each of the three potential moderators ("g", M-F, and A-R) trichotomizations were made. Table 7 presents the correlations between GPA and ACT for each of the prediction groups with respect to the three potential moderators. The difference between the predicted and both the overpredicted and underpredicted groups was significant ($p < .05$) when the trichotomization was based on M-F. Again, the differences were not significant in Sample B.

TABLE 5

CORRELATIONS BETWEEN GPA AND ACT FOR PREDICTION GROUPS IN SAMPLE B BASED ON CHI-SQUARE CLASSIFICATION

Problem	Prediction group					
	Overpredicted	N	Predicted	N	Underpredicted	N
1	.50	30	.56	98	.47	81
2	.55	85	.52	48	.43	76
3	.55	81	.59	49	.45	79
4	.52	101	.61	19	.48	89
5	.50	100	.61	29	.51	80
6	.47	78	.62	24	.53	107

TABLE 6

CORRELATIONS BETWEEN GPA AND ACT FOR FRACTIONATED GROUPS IN SAMPLE A

Variable[a]	Fractionated group	N	r
TA	$\frac{1}{3}$ low	72	.31
	$\frac{2}{3}$ high	137	.56
	$\frac{2}{3}$ low	138	.41
	$\frac{1}{3}$ high	71	.57
EA	$\frac{1}{3}$ low	67	.28
	$\frac{2}{3}$ high	142	.57
	$\frac{2}{3}$ low	137	.44
	$\frac{1}{3}$ high	72	.56

[a] Abbreviations: TA = Teacher Approval, EA = Education Acceptance. Correlations consecutively underlined are significantly different from each other at $p < .05$.

DISCUSSION

This article concerns the use of discriminant analysis for the systematic identification of moderator variables in an MPGV model. In addition, results from MPGV are compared with those of simple algebraic and absolute difference techniques. The results of this research did not support the use of any of the moderator variable techniques because of the problems and inconsistencies within each method.

Multipredictable Group Validation Model

The variables discriminating the groups in Sample A were not effective in Sample B. The ineffectiveness of discriminant analysis in this situation is due to several problems. First, the discriminant weights obtained in the initial MANOVAs were not different enough to lend themselves to an eyeball analysis for extracting significant contributors. Hence, MANOCOVA was necessary; the selection of the potential moderator variables was based on the a priori notion that the largest weights indicate the most important variables. It is possible that the selection of a different set of variables; for example, those that are significant in a univariate analysis of variance, would also provide a significant main effect. The necessity for using MANOCOVA limits the facility of using discriminant analysis to identify moderators.

Second, a significant discriminant function does not mean that all levels within a factor are discriminable. It is possible that the means for one set of levels are similar as are the means for a second set of levels; however, the first set of levels is different from the second set. Discriminant analysis does not necessarily provide a discrimination between all levels, but distinguishes between sets of levels.

Overlap or similarity of levels necessitated modification of the proposed procedure. For most problems studied, a discrimination between all levels of the prediction group was difficult. As a result, levels within a prediction group were combined. For some problems, it was further necessary to combine to the extent that the three basic prediction groups (overpredicted, predicted, and underpredicted) could not be maintained. The inability to maintain all levels and/or groups was due to the small amount of between group variance accounted for by the moderators. For the six problems, the proportion of between group variance accounted for by the set of moderators was .09, .12, .05, .14, .19, and .16, respectively.

Third, if two variables are highly correlated, they provide redundant information for discriminating among the groups. Thus, if a set of subtests are highly and significantly correlated, only one need be retained in the analysis. The procedure and interpretation would thus be facilitated.

Fourth is the problem of misclassification. The Index of Differentiation, a relatively crude estimate of classification, provided largely non-

TABLE 7

CORRELATIONS BETWEEN GPA AND ACT FOR PREDICTION GROUPS IN SAMPLE A BASED ON $\pm D$ CORRELATION WITH MODERATOR

Variable[a]	Prediction group	N	r
"g"	Overpredicted	62	.49
	Predicted	91	.42
	Underpredicted	56	.59
M-F	Overpredicted	53	.35
	Predicted	107	.66*
	Underpredicted	49	.49
A-R	Overpredicted	67	.42
	Predicted	84	.39
	Underpredicted	58	.35

[a] Abbreviations: "g" = Culture Fair Intelligence Test, M-F = Masculinity-Femininity, A-R = Analysis of Relationships.

* Predicted is significantly different from both overpredicted and underpredicted at $p < .05$.

ignificant associations. The chi-square tech-ique, derived from a discriminant analysis ystem, maximizes the probability of correct lassification. With both techniques, there was onsiderable misclassification as evidenced by the lack of many significant tetrachoric correlation coefficients. Examination of the contingency tables revealed that there was a tendency to place less than the actually observed number of *S*s into the predicted group.

Sample size is a related problem. Sample sizes of the prediction groups depend on three factors: (*a*) validity coefficient for the entire sample, (*b*) $\pm D$ interval for the predicted group, and (*c*) number of levels within the overpredicted and underpredicted groups. The higher the validity coefficient for the total sample and the greater the $\pm D$ interval for the predicted group, the more unequal are the prediction group sample sizes. In addition, if more than one level of underpredicted or overpredicted is used, the sample sizes of the levels within the unpredicted groups, especially the more extreme levels, become smaller. With relatively small and unequal sample sizes among groups, the test for significance between coefficients loses power.

Absolute Difference Technique

EA and TA were identified as moderators in Sample A. Results indicated that *S*s were more predictable if they had relatively higher acceptance of educational objectives and higher approval of their teachers' methods and classroom behaviors. The results did not hold up in cross-validation in Sample B, thus supporting the need for verifying moderator results prior to making decisions (Banas & Moore, 1968).

Algebraic Difference Technique

M-F, "g", and A-R were identified through their correlation with $\pm D$ as possible moderators in Sample A. The only variable that provided differential validities was M-F. The failure of "g" and A-R to yield differential patterns of validity may be due to restriction of range. Moderator variables are usually uncorrelated with predictors or criteria. In this study, M-F was uncorrelated with GPA and ACT, whereas "g" was correlated with GPA ($r = .23, p < .01$) and ACT ($r = .38, p < .01$), and A-R was correlated with GPA ($r = .42, p < .01$) and ACT ($r = .71, p < .01$). Consequently, subgrouping on the latter two potential moderator variables was equivalent to restricting the range on both the predictor and the criterion, resulting in lower validity coefficients.

With regard to M-F, differential validities appeared in Sample A. For those *S*s with normal masculine-feminine tendencies the validity coefficient was highest. The problem of cross-validation is evident in this technique as the $/D/$ technique.

Implications

The data of this research do not confirm the use of discriminant analysis to identify moderators in an MPGV model. Likewise, the data do not support the use of the algebraic difference or absolute difference (Ghiselli, 1956, 1960b) techniques. However, a consideration of the problems of discriminant analysis and those of cross-validation suggests further research.

First, future research should be conducted on larger sample sizes, thus limiting the effect of sampling error. Second, to facilitate procedure and interpretation, no more than one level within a prediction group is necessary.

Third, when potential moderator variables in a matrix are significantly and highly intercorrelated, only one variable of the set needs to be employed in the discriminant analysis. If the number of variables is extremely large, factor scores derived from a factor analysis can be used as the dependent variables in the discriminant analysis.

Fourth, since some evidence does exist that effective moderator variables may be uncorrelated with predictors and criteria, all potential moderator variables highly correlated with the predictor and/or criterion can be eliminated prior to subsequent investigation in a discriminant analysis. Also, the effect of a potential moderator as an independent predictor should be examined.

Fifth, once the moderators are determined, differential patterns of validity can be ascertained by considering the variables independently, one at a time. Thus, a simple analysis would be possible without interaction effects. Joint moderators were not found in academic situations in studies by Zedeck, Cranny, Vale, and Smith (1969) and Stricker (1966). In addition, the procedure would be easier since raw scores would be used in classification as opposed to discriminant scores.

Sixth, different classification procedures from the ones in this study might be tried. In addition, using more than one discriminant function for classification may account for more of the variance and result in better outcomes.

The use of a technique such as discriminant analysis in conjunction with MPGV does have some advantages. First, if variables could be found to identify the prediction groups, then greater emphasis in the future will be placed on reliability of instruments. Second, maintaining differences among unpredictable individuals enhances psychological understanding and leads to different priorities of selection. For example, suppose two individuals had the same acceptable predictor score but different moderator scores. On the basis of moderator scores, if one individual was placed in the underpredicted group and the other in the predicted group, then the MPGV concept suggests hiring the former before the latter. In essence, MPGV reduces the amount of misses more than does the absolute difference technique.

Finally, a systematic technique for selection of moderator variables would allow us to examine larger masses of data and select those variables which serve as moderators. A systematic technique is more efficient than intuition for discovering or developing moderator variables.

REFERENCES

ABRAHAMS, N. M. Off-quadrant comment. *Journal of Applied Psychology*, 1969, **53**, 66–68.

AMERICAN COLLEGE TESTING PROGRAM. *American college test*. Iowa City: American College Testing Program, 1965.

BANAS, P. A. An investigation of transsituational moderators. Unpublished doctoral dissertation, University of Minnesota, 1964.

BANAS, P. A., & MOORE, L. Moderating effect of job satisfaction and evaluating moderator effectiveness. Paper presented at Midwestern Psychological Association, Chicago, May 1968.

BELL, H. M. *Bell adjustment inventory*. Palo Alto: Consulting Psychologists Press, 1962.

BROWN, W. F., & HOLTZMAN, W. H. *Survey of study habits and attitudes*. New York: Psychological Corporation, 1965.

CATTELL, R. B. *Culture fair intelligence test*. San Diego: Educational & Industrial Testing Service, 1963.

COOLEY, W. W., & LOHNES, P. R. *Multivariate procedures for the behavioral sciences*. New York: Wiley, 1962.

DUNNETTE, M. D. A modified model for test validation and selection research. *Journal of Applied Psychology*, 1963, **47**, 317–323.

ENGLAND, G. W. *Weighted application blanks*. Dubuque, Iowa: Brown, 1961.

EYSENCK, H. J. *Maudsley personality inventory*. San Diego: Educational & Industrial Testing Service, 1962.

FREDERIKSEN, N., & MELVILLE, S. D. Differential predictability in the use of test scores. *Educational and Psychological Measurement*, 1954, **14**, 647–656.

FRENCH, J. W. A machine search for moderator variables in massive data. Princeton: Educational Testing Service, 1961.

GAYLORD, R. H., & CARROLL, J. B. A general approach to the problem of the population control variable, *American Psychologist*, 1948, **3**, 310. (Abstract)

GHISELLI, E. E. Differentiation of individuals in terms of their predictability. *Journal of Applied Psychology*. 1956, **40**, 374–377.

GHISELLI, E. E. *Analysis of relationships*. Palo Alto: Consulting Psychologists Press, 1960. (a)

GHISELLI, E. E. The prediction of predictability. *Educational and Psychological Measurement*, 1960, **20**, 3–8. (b)

GROOMS, R. R., & ENDLER, N. S. The effect of anxiety on academic achievement. *Journal of Educational Psychology*, 1960, **51**, 299–304.

HOBERT, R., & DUNNETTE, M. D. Development of moderator variables to enhance the prediction of managerial effectiveness. *Journal of Applied Psycology*, 1967, **51**, 50–64.

LYKKEN, D. T., & ROSE, R. Psychological prediction from actuarial tables. *Journal of Clinical Psychology*, 1963, **19**, 139–151.

MCNEMAR, Q. Moderation of a moderator technique. *Journal of Applied Psychology*, 1969, **53**, 69–72.

SAUNDERS, D. R. The "moderator variable" as a useful tool in prediction. *Proceedings of the 1954 Invitational Conference on Testing Problems*. Princeton: Educational Testing Service, 1955.

SAUNDERS, D. R. Moderator variables in prediction. *Educational and Psychological Measurement*, 1956, **16**, 209–222.

STRICKER, L. J. Compulsivity as a moderator variable: A replication and extension. *Journal of Applied Psychology*, 1966, **50**, 331–335.

TOOPS, H. A. A research utopia in industrial psychology. *Personnel Psychology*, 1959, **12**, 189–225.

ZEDECK, S., CRANNY, C. J., VALE, C. A., & SMITH, P. C Comparison of "joint moderators" in three predictio techniques. *Journal of Applied Psychology*, 1971, **55** 234–240.

ZUCKERMAN, M., & LUBIN, B. *Multiple affect adjecti check list*. San Diego: Educational & Industri Testing Service, 1965.

(Received March 23, 1970)

TABLE 2

Means and Standard Deviations of General Aptitude Test Battery Aptitudes by Age and Occupation

Warner level	Age group 25–35	36–40	41–45	46–50	51–55	56–83
Intelligence						
1–2						
$\bar{X}$	125.9	123.9	123.3	120.6	118.0	111.9
SD	14.1	17.3	15.2	14.2	16.1	15.1
3						
$\bar{X}$	118.2	119.3	115.8	115.3	111.4	109.1
SD	16.7	13.7	14.0	14.5	15.5	18.6
4						
$\bar{X}$	107.5	108.5	104.8	105.2	108.1	99.4
SD	15.0	15.1	14.3	16.3	15.2	20.1
5–7						
$\bar{X}$	104.4	102.8	105.2	101.8	100.7	98.0
SD	16.0	14.6	14.2	13.7	16.1	12.0
Verbal Aptitude						
1–2						
$\bar{X}$	121.5	120.8	120.8	116.0	116.2	112.7
SD	14.3	17.1	15.7	15.1	15.2	16.5
3						
$\bar{X}$	116.6	115.2	112.3	114.5	111.6	109.8
SD	16.9	14.1	13.5	14.6	16.4	16.1
4						
$\bar{X}$	104.2	107.2	104.6	107.4	110.9	100.1
SD	13.4	14.3	12.2	16.6	11.9	19.4
5–7						
$\bar{X}$	102.5	101.3	104.3	103.4	101.8	102.6
SD	15.4	14.5	12.7	12.8	15.4	13.8
Numerical Aptitude						
1–2						
$\bar{X}$	126.4	122.2	123.2	102.3	117.6	114.9
SD	15.4	16.1	16.1	15.0	16.1	16.1
3						
$\bar{X}$	118.5	118.7	115.7	112.6	110.9	108.5
SD	16.3	15.5	16.0	19.1	16.2	20.2
4						
$\bar{X}$	104.9	106.5	105.0	103.8	108.5	98.1
SD	14.6	14.3	13.4	16.4	18.4	18.8
5–7						
$\bar{X}$	102.8	102.5	104.3	98.2	99.2	94.7
SD	16.0	14.3	12.8	13.2	15.3	14.1
Spatial Aptitude						
1–2						
$\bar{X}$	114.6	113.7	111.3	107.9	105.2	96.3
SD	17.2	18.9	16.4	16.3	16.2	14.4
Spatial Aptitude (continued)						
3						
$\bar{X}$	108.7	110.8	109.6	107.5	98.0	96.2
SD	17.1	17.5	15.3	14.9	15.2	15.9
4						
$\bar{X}$	105.7	105.5	98.5	99.2	93.3	90.8
SD	14.5	18.1	18.1	16.4	15.1	18.5
5–7						
$\bar{X}$	106.3	99.2	103.9	100.6	94.2	91.3
SD	15.1	15.7	14.1	15.6	16.9	13.3
Form Perception						
1–2						
$\bar{X}$	117.7	115.5	113.4	105.6	100.5	91.8
SD	15.5	17.6	18.0	15.4	16.0	16.4
3						
$\bar{X}$	112.7	109.8	107.1	104.3	97.7	90.3
SD	19.0	20.0	19.3	19.8	19.1	16.9
4						
$\bar{X}$	107.8	109.4	101.0	94.7	96.5	87.8
SD	18.0	16.9	16.3	17.5	12.9	23.0
5–7						
$\bar{X}$	104.5	98.7	104.0	96.6	89.8	87.8
SD	20.7	14.1	16.2	15.7	13.9	18.7
Clerical Perception						
1–2						
$\bar{X}$	124.9	125.8	124.5	120.1	118.8	115.8
SD	16.2	16.1	15.2	16.0	15.7	13.8
3						
$\bar{X}$	123.8	123.6	117.4	118.6	113.7	113.4
SD	17.9	16.8	15.4	16.9	15.7	14.1
4						
$\bar{X}$	114.4	114.8	109.3	113.1	109.7	103.9
SD	13.8	12.7	12.3	15.4	17.0	18.5
5–7						
$\bar{X}$	110.9	108.1	114.5	106.2	105.4	103.3
SD	16.8	13.1	13.1	12.8	13.0	13.1
Motor Coordination						
1–2						
$\bar{X}$	118.2	117.8	112.8	115.2	111.3	107.8
SD	14.7	17.8	16.6	16.9	17.6	20.5
3						
$\bar{X}$	117.9	116.6	111.9	113.0	108.5	100.5
SD	18.5	16.2	16.7	21.4	15.4	19.5

TABLE 2—(*Continued*)

Warner level	Age group 25–35	36–40	41–45	46–50	51–55	56–83
Motor Coordination (continued)						
4						
$\bar{X}$	105.6	106.4	107.5	107.2	105.1	98.2
SD	18.5	16.3	18.4	15.4	20.6	22.1
5–7						
$\bar{X}$	104.9	104.0	107.7	100.1	98.8	95.2
SD	19.9	13.0	12.7	12.7	16.8	17.5
Finger Dexterity						
1–2						
$\bar{X}$	98.8	89.3	89.6	85.6	77.1	70.9
SD	20.7	17.8	19.4	18.9	19.0	20.0
3						
$\bar{X}$	92.9	91.2	86.0	83.4	76.7	65.5
SD	21.0	19.0	20.3	20.3	20.7	23.1
4						
$\bar{X}$	95.6	89.8	86.8	79.5	75.8	73.5
SD	16.6	16.1	17.3	17.4	19.9	24.8
5–7						
$\bar{X}$	87.5	85.1	86.2	81.0	70.0	62.3
SD	16.8	18.2	18.1	18.8	13.0	19.0
Manual Dexterity						
1–2						
$\bar{X}$	107.2	102.5	98.4	93.6	92.4	80.1
SD	23.1	18.4	19.5	17.0	18.9	18.4
3						
$\bar{X}$	104.8	108.3	99.2	97.6	96.4	82.4
SD	20.1	20.1	21.8	22.0	20.1	19.2
4						
$\bar{X}$	106.3	103.5	99.9	94.8	88.3	89.3
SD	20.3	17.3	17.3	20.9	22.8	22.8
5–7						
$\bar{X}$	103.9	98.9	101.3	99.6	86.2	76.4
SD	19.9	18.4	15.7	17.0	15.7	25.4

Warner Levels 6 and 7 are 21% and 10% of the group, respectively, in the present study.

Nine two-way analyses of variance were conducted on the aptitude scores with age and SES as factors. (A discussion of similar analyses performed on the 12 GATB subtests are described elsewhere [Fozard & Nuttall, 1970; Nuttall & Fozard, 1970].) The unweighted cell means procedure was employed for each. Results of the analyses showed statistically significant age effects for all aptitude scores. Verbal Aptitude was significant at the .05 level; the others were significant at the .001 level. In all cases, mean scores declined with age. The effect of SES was statistically significant for eight of the nine aptitudes. Manual Dexterity did not change significantly with SES. Finger Dexterity was significant at the .01 level, while the other seven aptitudes were all significantly affected by SES at the .001 level. In all cases the men from lower SES levels had lower scores for the eight aptitudes significantly affected by SES.

For none of the nine aptitude scores was there a statistically significant interaction between age and SES.

The percentage of variance accounted for by the main effects of age and SES in the nine analyses of variance was computed based on the procedure described by Hays (1963). The results are shown in Figure 1. The percentage accounted for by SES in each aptitude is on the horizontal axis, that for age on the vertical axis. The magnitudes of the values in Figure 1 reflect the fact that differences in SES are reflected most in Numerical Aptitude, Intelligence, Verbal Aptitude, and Clerical Perception. In contrast to the results of the "four state" study (United States Department of Labor, 1967, p. 225), the effect of age on Ver-

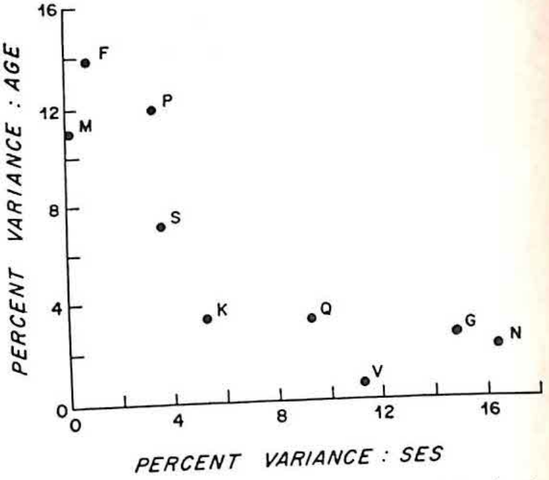


FIG. 1. Percentage of variance accounted for in nine GATB aptitudes by differences in four Warner level (SES) and six age classifications. (Percentages of variance were calculated assuming a fixed factor analysis of variance. Abbreviations: F = Finger Dexterity, M = Manual Dexterity, P = Form Perception, S = Spatial Aptitude, K = Motor Coordination, Q = Clerical Perception, V = Verbal Aptitude, G = Intelligence, and N = Numerical Aptitude.)

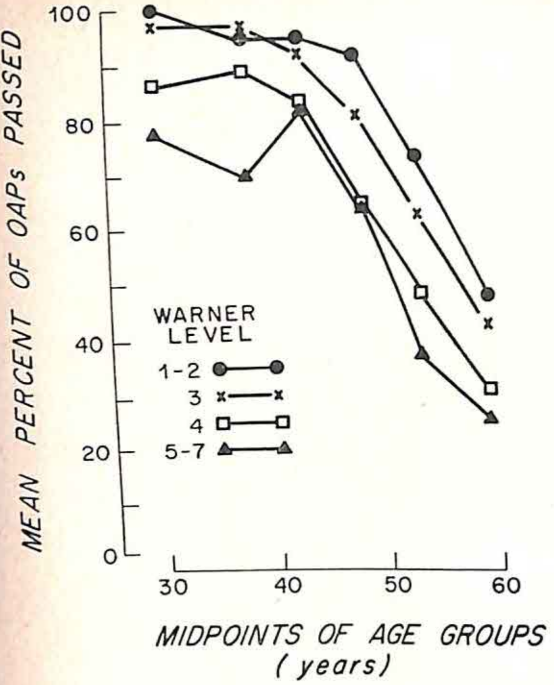


FIG. 2. Percentage of 36 Occupational Aptitude Patterns for which the average of men in each of the combinations of age and SES qualify.

bal Aptitude was statistically significant in the present study.

The hypothesis of curvilinearity in the effect of age on Intelligence, Verbal Aptitude, Numerical Aptitude, and Spatial Aptitude (Hirt, 1959) was examined. Trend analyses, using Gaito's (1965) procedure for unequal numbers of cases and unequal intervals, were performed for the age differences within each Warner Level separately. The results of the 36 trend analyses are summarized elsewhere (see Footnote 3).

In the present sample Warner Levels 5–7 are most similar to the level of occupations represented in Hirt's (1959) study. In contrast to Hirt's results, in the present study *none* of the quadratic trends in that group was significant at the .05 level. In Warner Level 4, however, the quadratic trends were significant for all aptitudes except for Clerical Perception. For the Warner Level 3 group only Spatial Aptitude and Manual Dexterity had significant quadratic trends. At Warner Level 1–2 Intelligence, Spatial Aptitude, and Form Perception had significant quadratic trends.

Examination of Table 2 shows that most of the significant quadratic trends occurred because the differences between the two or three oldest groups were larger than those between the younger groups. The exceptions are the quadratic trends for Intelligence, Verbal Aptitude, and Numerical Aptitude in Warner Level 4, in which the means for the 51–55-yr. age group were higher or almost as high as in any of the younger groups.

The linear trends were statistically significant in 27 out of 36 cases. Most consistent linear trends in all SES groups were those for the psychomotor aptitudes, Finger Dexterity and Manual Dexterity.

The procedure used by Heron and Chown (1967) to assess differences in the age trends of the two extreme SES classifications was applied to the age differences for Warner Levels 1–2 and 5–7. The differences between the means of these two Warner level groups in each of the six age groups were found for the nine aptitudes, and the rank order of correlation between these differences and age were computed. Only for Intelligence and Verbal Aptitude were the rank order coefficients significant beyond the .05 level. The significant correlations occurred because the differences between the two extreme groups grew *smaller* as age increased. The age decline was greater in the Warner Level 1–2 groups (see Table 2). These results are the opposite of those found by Heron and Chown (1967).

Occupational Aptitude Patterns. The mean score on the aptitudes for each of the 24 groups was compared to the cutoff scores for each of the 36 OAPs (United States Department of Labor, 1962). The percentage of OAPs for which these mean scores qualify is indicated in Figure 2. Clear differences are observed between the SES levels, especially between Warner Levels 1–2 and 3, as contrasted to Warner Levels 4 and 5–7. Age, however, seemed to have a stronger effect than SES, with all four SES groups dropping precipitously with advancing age. The young Warner Level 1–2 group qualified, on the average, for all 36 OAPs, but the oldest group for this SES level qualified on the average for only 17 OAPs. Similarly, the lowest SES group qualified for 28 OAPs on the basis of their average scores at the youngest age, but only for 9 OAPs at the oldest age. Clearly, the older worker in all SES groups is severely handicapped when compared to younger men on the basis of OAP cutoff scores.

DISCUSSION

The major results of the study may be summarized as follows: (*a*) Both age and SES affect average performance on the GATB aptitude scores and the average percentage of OAPs passed. (*b*) There were no statistically significant interactions between the effects of the two variables when all Warner Levels were considered. When the data of only the two extreme SES groups were used, however, in GATB aptitude of Intelligence and Verbal Aptitude, there was a statistically significant tendency for the differences in the scores of the high and low SES groups to become *smaller* with age. (*c*) In general, aptitude scores which were most influenced by age were least influenced by SES and vice versa.

There was little evidence in the present data that the curvilinear trend in age differences for Intelligence, Verbal Aptitude, Spatial Aptitude, and Numerical Aptitude found by Hirt (1959) generalizes across SES classifications. The occupational levels of his sample were only poorly represented in the present group so that a direct comparison was not possible.

There were many statistically significant curvilinear trends for age differences in the GATB aptitude scores. Except for Warner Level 4, these trends could be attributed to differences in the amount of decline between older and younger age groups, that is, a threshold effect.

The exceptions in the Warner Level 4 groups occurred because performance in the 51–55-yr.-old group was high in the aptitude scores most sensitive to SES. One hypothesis to explain this effect is that the initial occupation of *S*s in this group may have been of a lower status than that suggested by their aptitudes. Men in this age group typically held their first jobs during the depression years of the 1930s, when there were fewer jobs available in the higher SES categories. In a depressed job market, a given pattern of abilities would lead to a lower occupational attainment, hence lower SES, than would be the case in an expanding job market. Many men of the depression era would then be "frozen" in an occupation below what their abilities would have allowed them to achieve in better economic times. Some support for this hypothesis comes from Table 1 which shows that the educational level in the 51–55-yr.-old group is somewhat higher than average for Warner Level 4.

The results of all the analyses of the present data show unequivocally that there is no statistically significant interaction between the main effects of age and SES. Thus, there is no evidence in the present data analysis to support the idea that age decrements in cognitive functioning decline less in high- than in low-SES groups.

Application of Heron and Chown's (1967, pp. 104–106) procedure for assessing differential declines in extreme SES groups showed that the amount of decline in Intelligence and Verbal Aptitude was different for Warner Levels 1–2 and 5–7. In contrast to Heron and Chown's results, however, which showed the largest differences between the two SES groups at the oldest ages, the present study found the two SES groups becoming more similar in Intelligence and Numerical Aptitude with increasing age.

Even allowing for differences in tests and systems of SES classifications in the two samples, it would not be expected that the age trends would be directly opposite, particularly in Intelligence, which is the best unitary measure of general intelligence available in the GATB (United States Department of Labor, 1967, Ch. 14). One explanation of the discrepancy is that the educational differences in the two extreme SES groups were greater among the older men than among the younger men in the British sample. Alternatively, one could argue that the lowest SES groups were not well represented in the present study, particularly in the oldest age groups. As a result, the observed age differences between extreme SES groups would not be as large in the present sample as in the general population.

An additional possibility is that the differential age effect in high and low SES groups observed by Heron and Chown (1967) is peculiar to the Progressive Matrices Test. The GATB Aptitude of Intelligence, and also Numerical Aptitude, measure more the "crystallized" intelligence that the "fluid" intelligence as indicated by the Progressive Matrices

urnal of Applied Psychology
71, Vol. 55, No. 4, 380–383

METHOD FOR DETECTING RANDOM ANSWERS ON PERSONALITY QUESTIONNAIRES

JERRY W. O'DELL[1]

Eastern Michigan University

Several strategies for detection of randomly answered protocols for the Sixteen Personality Factor Questionnaire, Form A, 1967 edition were developed and evaluated. Cattell's Sabotage index predicted poorly, indeed in the wrong direction, between normal and randomly answered sheets ($t = 2.10$, $p = .05$). Random and Carelessness indexes were derived by the author for the same purpose and predicted very well, providing *t*s of 15.02 and 12.67 (significant at well beyond the .001 level), respectively, between random and normal sheets. More importantly, both indexes, while correctly selecting 86% and 88% of the random answer sheets, respectively, incorrectly picked out only about 6% of the normal group. Thus, while detection of random protocols is rather accurate, few normal people would be offended by having their accuracy questionned. Since the design of the study included a form of cross-validation, it is suggested that these indexes might well be of great practical use.

Randomly filled-in answer sheets pose nightmare-like problems for any person putting psychological tests into practical use. Recent rapid development of automated scoring and interpretation systems for tests (see Butcher, 1969) is compounding the problem by leaps and bounds. The emphasis of these systems is more and more that of producing a report indistinguishable from that produced by a human psychologist, and there seems little doubt that really elegant (although not necessarily valid) systems will be available in the very near future. These will be capable of describing everything from probable work habits to clinical syndromes. This development will make it possible for a personnel worker or counselor to obtain a report looking almost exactly like the routine production of an expert test interpreter with little more cost or effort than is required for simple scoring, and on tens of thousands of persons. These elegant programs unfortunately have no way of distinguishing a randomly answered test from one completed with care. They can and will generate beautifully believable reports on a meaningless protocol. It should be evident that if such great numbers of persons are tested, a discouraging number of absurd reports is going to be developed on randomly answered sheets. The author, in attempting to develop an automated interpretation system for the 1967 Sixteen Personality Factor Questionnaire, Form A, (16PF; Cattell, Eber, & Tatsuoka, 1970) ran up against the problem of random answering so frequently that it became clear that a means would have to be developed to solve it.

Although the basic problem is that of picking out as many of the random profiles as possible, one error in particular cannot be tolerated, namely, calling a sheet randomly answered when it has in fact been filled out in good faith. Such an error leads to the difficult ethical problem of having to incorrectly question someone's honesty. A high proportion of errors of this sort could lead to chaos in any organization, and any method used should make as few errors as possible in this regard.

Over the years there have been many attempts at developing indexes of careless answering, none of them apparently having been demonstrated to be successful. For the 16PF, Cattell himself (Cattell, Eber, & Tatsuoka, 1970) has considered the problem of random answering, which he colorfully calls "sabotage." He has proposed a method for sabotage detection on purely theoretical grounds. The basic assumption of this procedure is that the score on one-half of a given 16PF scale should be predictable from the other half of the scale through regression techniques. For each scale, the difference between the actual and predicted score is obtained, each difference is squared, and the squared differences are summed across all 16 scales to provide the sabotage index. It is said, on purely theoretical grounds, that one may be 99% sure of sabotage if the score exceeds 102, and 95% sure if it exceeds 83. To the best of this author's knowledge, this index has never been systematically tested, and such a check clearly needs to be done.

---

[1] Requests for reprints should be sent to Jerry W. O'Dell, Eastern Michigan University, Ypsilanti, Michigan 48197.

Preliminary work with Cattell's index suggested that it was not effective, and thus it seemed necessary to try other methods of detection. One method, used in this article, employs extreme item alternative response frequencies. The frequency of response to each alternative of each question is tallied in a group of persons motivated to answer accurately. Those alternatives that are chosen by very few people in the population, say by less than 10%, are then selected and put together to make up a "random" scale. Accurate answering of the questionnaire should yield a low score on this scale; careless responding should lead to a much higher, chance score. Such a scale may also be constructed from alternatives which virtually every person in a motivated group selects, say 90% or more. Here careful answerers should obtain a high score, with randomness being indicated by a lower, chance score. Variations in this approach have been used in the construction of the *V* score on the various Kuder Preference Records (Kuder, 1950, 1964) and in the Minnesota Multiphasic Personality Inventory *F* scale (Dahlstrom & Welsh, 1960). For some reason, however, there seems to be no systematic research reported on the efficacy of these methods in practice. This may well be due to an understandable tendency of workers to focus on the more glamorous topics of "faking good" and "faking bad."

In addition, it seemed desirable to measure some very specific and obviously deviant response sets, such as answering all the questions with the same alternative, leaving too many items blank, or having an inordinate number of multiple responses. The 16PF also contains three direct questions about accuracy of responding (Numbers 1, 2, and 187) that ask whether one has understood the directions, whether one is prepared to answer truthfully, and whether one is sure that he has not omitted any responses. Obviously, if an *S* shows doubt about any of these, there should be doubt about the whole record. It was the various indexes mentioned in this paragraph that were combined to provide the "carelessness" index mentioned subsequently.

The purpose of this article, then, is to evaluate the practical effectiveness of Cattell's Sabotage index, to construct and test a single scale for the detection of random answering, and, finally, to develop and check a summary index for as many indicators of careless answering as possible.

METHOD

Development of the original scales was done on a sample of 250 students in introductory psychology classes at Eastern Michigan University. Virtually all *Ss* were freshmen and sophomores between the ages of 17 and 20. The sample was well balanced with regard to sex. The 16PF was administered during a normal class period, and students were informed that they would be able to see the results of the test in private consultation with the instructor. The prospect of seeing the results engendered obvious interest in the test. This interest was essential to keep the number of carelessly answered sheets to a minimum. All scoring and data processing was done through the use of an OpScan scanner and an IBM 1130 computer.

Evaluation of Cattell's sabotage score was done through a computer subroutine, which, because of the complexity of the calculations, was checked many times against hand-computed results to insure accuracy.

The scale to detect random answering (hereafter referred to as random) was constructed by the method of selecting items with extreme item alternative response frequencies as was mentioned earlier. There are two basic ways of doing this. (*a*) An alternative may be selected for the scale if it is answered by a very small percentage of the sample. In this case, if selected by an individual, it would count one point toward the random score. (*b*) An alternative may be selected for the scale if it is selected by a very large percentage of the sample. Such an item would count one point toward the random score if it were *not* selected by a respondent. Note that on the usual true–false test, the two ways are essentially the same. For example, if a given item is answered true by 10% of the sample, it will of necessity be answered false by 90% of the sample.

The existence on the 16PF of three alternatives—A, B, and C—leads, however, to a troublesome ambiguity. Consider an item with alternative response frequencies of 2% A, 49% B, and 49% C. Such an item is an obvious candidate for the first strategy mentioned above and could clearly be made to contribute one point to the random score if it were answered A. On the other hand, to illustrate the ambiguity, consider an item with alternative response frequencies of 1% A, 12% B, and 87% C. Notice here that under the first strategy either Item A or Item B might be considered infrequent. But the question arises as to which of the two should be chosen or whether both should be picked. Fortunately, there is another way of looking at this last set of percentages, namely, that what was really happening was not so much that Items A or B were being rejected, but rather that C was being *preferred*. It makes better conceptual sense in such a case to use the second strategy, that of scoring the high frequency alternative if it is *not* selected by a testee. Thus, in the example, Alternative C, selected by 87% of the sample, would be scored if not answered, and the ambiguity about which of the other two alternatives to score is eliminated.

Both strategies were adopted in the construction of the present random scale. Specifically, an item was chosen for the scale if it were answered by 10% or less of the sample, assuming that the other alternatives were both frequently answered. If they were not both frequently answered, however, (i.e., if the second situation mentioned above arose), then the item was made to contribute one point to the random score if the frequently answered alternative was not chosen. In the latter case, it was initially planned to use only alternatives chosen by 90% or more of the sample, but it was found that this cutting point could not be held as rigidly as could the 10% point. The average percentage of choices, however, was 90.35 for all the items selected through the second strategy. The only serious deviation from this cutoff point was one item that had a popularity of only 80.4% but which still seemed effective.

Items selected for the random scale through the procedure mentioned were as follows: Numbers 38, 54, 77, 78, 80, 93, 102, 118, 119, 121, 143, 1*, 2*, 29*, and 112* added one point to the score if answered A. Numbers 12, 20, 23, 51, 52, 66, 87, 91, 109, 127, 28*, 53*, and 128* counted if answered B. Only three were keyed C, namely, 4, 73, and 116*. Items with asterisks counted if they were not answered. Items picked for this scale are distributed fairly evenly over the scales of the 16PF with the exception that 8 of the 31 were on the intelligence scale, B. Exclusion of these items unfortunately reduced the efficiency of the scale, but it is probably reasonable to hope that the 16PF will be used primarily with college students or persons who are college trained. Hence, these intelligence items should remain a fairly constant factor.

In detecting extreme response sets, simple cumulative tallies were made of the answers to the various alternatives. It was found that of the 250 *S*s no one selected 33 or fewer, or 100 or more A responses; no one selected 8 or fewer, or 100 or more B responses; and no one selected 29 or fewer, or 93 or more C responses. It was arbitrarily decided that more than five multiple responses or more than five blank responses would be considered excessive. Exceeding any of these limits was presumed to be evidence of carelessness.

To this point some 15 indexes of careless answering have been developed, and scanning all of these on all records proved to be a cumbersome chore. What was needed was a simple summary score that would signal trouble with the record at a glance. It was arbitrarily decided that such a score should be nearly zero in the absence of suspicious indicators, but should rise above 10 if trouble of any kind was sensed. Consequently, it was decided that detection of the really serious deviations—too few or too many As, Bs, or Cs; too many blanks; or too many multiple responses—should contribute 10 points to this summary score. It was originally intended to add 10 points to *S*s' score for answers other than "yes" to the three 16PF questions asking directly about the accuracy of responding, but the item counts showed that many people actually had doubts about their care in answering. To face this reality, an answer of "uncertain" was made to add only two points to the summary score, while an outright avowal of doubt added five points. Finally, since the average normal record yielded a random score of only 1.58, while an average carelessly answered record led to a score of 7.80, doubling the random score would result in a score of about 17 for randomly answered sheets, which is in the desired range. The cumulative combination of these scores went to make up the "carelessness" index.

The scales were checked out on two groups of answer sheets. One group (hereafter referred to as normal) consisted of 173 16PFs that had been collected over the years on upper-class and graduate psychology students. This group had no overlap with the original sample of 250 used to develop the scales. The second group, referred to as the random group, consisted of 59 randomly answered sheets. Seventeen of these were made up to show certain definite response aberrations. The remaining 42 were filled out randomly by students with no 16PF booklet at hand.

RESULTS

The Sabotage index did not fare well in this study. Using the cutting score of 83, it indicated that 110 of the 173 persons in the normal group were inattentive (63.5%), while it successfully picked out only 30 of the 59 actually random sheets (50.8%). Corresponding figures for a cutting score of 102 are 70 normals misclassified as random (40.5%), with only 17 (28.8%) of the randoms correctly selected. For the normal group the mean sabotage score was 97.98 with a standard deviation of 32.08. For the random group the mean score was 88.50 with a standard deviation of 29.23. The difference between these means results in a t of 2.10 ($df = 230$, $p = .05$) but in the wrong direction. Assuming these findings to be generally true, this number of misclassifications would be simply unacceptable.

The random score was much more successful. Taking a score of five or higher as evidence of random answering, 10 normals out of 173 (5.8%) were incorrectly called random, while 51 out of 59 (86.4%) of the true randoms were selected. For the normal group, the mean score on the random scale was 1.58, with a standard deviation of 1.68; corresponding figures for the random group were 7.80 and 3.04. The t between these means was 15.02, again highly significant.

The Carelessness index had very nearly the same degree of success. Considering a score of 16 or higher as evidence of random answering, 10 of the 173 normal protocols (5.8%) were called random, and 52 of the 59 random protocols were correctly selected (88.1%). The mean for the normal group was 5.64, with a standard deviation of 6.12; for the random group the corresponding figures were 29.17 and 13.81. The *t* between these two means was 12.67, again most significant.

DISCUSSION

The random and carelessness scales appear to be promising indicators of random answering, especially since they have been checked out on a group other than that on which they were derived, and, hence, to a degree have been cross-validated. Moreover, it is likely that some of the 173 normal protocols were in fact randomly answered, so that their efficiency may be slightly underestimated. And while the significance tests of the means are of interest, probably the more important consideration is that relatively few people who were honest in answering would be incorrectly accused of being careless (5.8%). This is not ideal, but perhaps sufficiently accurate that a human or computer scorer could state with some definiteness that a given answer sheet was suspect. The most dramatic demonstration of the effectiveness of the carelessness index occurred when the author accidentally misplaced one of the OpScan output cards used for computer scoring in another sample of cases. This had the effect of making all the protocols after this point random, and the carelessness scores derived after this point jumped from virtually all zero to a rather steady value in the 30s.

With regard to the relative usefulness of the random and carelessness scores, it is the author's feeling that the latter is the method of choice, since it checks on a much wider range of problems. In addition, if computer scoring is used, the machine may easily be made to print out explicit comments as to what the particular difficulties are, thus bringing them to the direct attention of the user.

Cattell's "Sabotage" index, from the results of this study, seems of little use. It may be unfair, however, to condemn it outright. Cattell frequently warns that all forms—A–E—of the 16PF should be given to obtain maximum discrimination. In using only Form A, this warning was not heeded. It does seem unfair to base this score on halves of scales containing only 10 or 13 items. At the same time, in most selection situations only one form of the 16PF is given. The great computational labor involved in this index (at least an hour of frustrating work with a calculator) would seem to make its use unwise until much better evidence can be obtained of its effectiveness. This is especially true in view of the fact that the random scale developed in this article can be scored in about 1 min. with far less chance for error.

REFERENCES

BUTCHER, J. N. (Ed.) *MMPI: Research developments and clinical applications.* New York: McGraw-Hill, 1969.

CATTELL, R. B., EBER, H. W., & TATSUOKA, M. M. *Handbook for the 16PF.* Champaign, Ill.: Institute for Personality and Ability Testing, 1970.

DAHLSTROM, W. G., & WELSH, G. S. *An MMPI handbook: A guide to use in clinical practice and research.* Minneapolis: University of Minnesota Press, 1960.

KUDER, G. F. Identifying the faker. *Personnel Psychology*, 1950, **3**, 155–167.

KUDER, G. F. *Kuder E general interest survey manual.* Chicago: Science Research Associates, 1964.

(Received April 2, 1970)

urnal of Applied Psychology
71, Vol. 55, No. 4, 384–388

MULTIPLE–INDICANT APPROACH FOR STUDYING NEW PRODUCT ADOPTERS[1]

JACOB JACOBY[2]

Purdue University

Aspects of the traditional approach to identifying and studying innovators (i.e., select the first $2\frac{1}{2}\%$ or 10% to try a single new product and compare characteristics of these individuals to later triers and nontriers) were criticized. In contrast, an innovation proneness index, ranging from 0 to 15, was derived based on responses to experimental stimuli presented for 15 product classes. Sixty female undergraduates, confronted by one new and four old items in each product class, role-played purchase selections. Using this index, the hypothesis that low dogmatic individuals would make significantly more innovative responses than would high dogmatics was confirmed.

The vast majority, perhaps 90% (cf. Rogers & Stanfield, 1968), of the thousands of new products introduced annually end up being rejected by the consumer. Given the high costs involved in the research, development, and introduction of new products, identifying the innovation prone individual (i.e., the consumer most likely to be among the first to try a new product), especially those factors which contribute to his being an innovator, assumes tremendous practical importance. Not only do such individuals provide a toehold for new products, but they also seem to play an integral part in the subsequent diffusion of the innovation through the population at large (cf. Engel, Kegerreis, & Blackwell, 1969; Robertson, 1968; Rogers, 1962; and Zaltman, 1965, p. 31).

Generalizations from research conducted thus far (cf. Bohlen, Coughenour, Lionberger, Moe, & Rogers, 1968; Coleman, 1960; King, 1963; Robertson, 1968; and Rogers, 1962) indicate that innovators tend to have higher income and status and be younger, more mobile, educated, literate, and privileged (i.e., have more discretionary income relative to their peers). Personality characteristics most often cited are venturesomeness (i.e., possessing favorable attitudes toward trying new ideas and products) and greater risk-taking tendencies than noninnovators.

While these findings are interesting and suggestive, several factors limit their generalizability. Consider the typical study. As a rule, investigators select a single product, usually a luxury item when studied (e.g., touch-tone telephones, color TV, electric toothbrushes, electric knives, stereo phonographs, auto diagnostic centers), and arbitrarily define innovators to be the first $2\frac{1}{2}\%$ or 10% of all potential adopters to purchase the item. All remaining potential (and actual) adopters are considered to be noninnovators. Characteristics of individuals in the two groups are then compared, often on the basis of raw frequencies or percentages, and conclusions drawn.

Limitations inherent in such a procedure are as follows. First, obtaining only one indication of an individual's tendency to innovate (i.e., using early trial of a single product) severely limits one's ability to generalize about the characteristics of individuals who tend to be innovators in general, regardless of the specific product class. Second, given the luxury-type products typically selected, it is not surprising that innovators tend to have higher incomes and be more privileged. The finding that innovators take greater risks may simply reflect their greater financial ability to afford such risks, rather than being a function of some underlying personality trait. Indeed, such factors as income, education, and occupation appear to have little relation to early adoption when the innovation is a new brand of a frequently purchased, low priced product such as coffee (Frank, Massy, & Morrison, 1964).

Third, a trial versus nontrial criterion ignores the fact that new product adoption represents a decision-making process of several stages. According to the most generally accepted model (Rogers, 1962), trial is preceded by stages of awareness, interest, and evaluation, and the decision to continue full use of an item (i.e., adoption) is a postrial phenomenon. Treating adoption dichotomously by comparing characteristics of triers and nontriers is of limited value in understanding the psychological dynamics preceding and underlying innovation. Last, identifying correlates of innovativeness, expecially when these are not predicted in advance, provides minimal information regarding cause-effect relationships.

---

[1] The author would like to express his appreciation to Ann P. Reed and Sharon L. Towles for their assistance in developing this instrument and serving as *E*s.

[2] Requests for reprints should be sent to Jacob Jacoby, Department of Psychology, Purdue University, Lafayette, Indiana 47907.

In contrast, making adequate generalizations requires obtaining several indexes of innovation proneness across a variety of commonly purchased, low priced, necessity-type items. Another desideratum is experimentation based on a priori hypotheses. The present study is cast in this mold.

Despite the generally disappointing results of studies trying to link personality to consumer behavior (cf. Engel, Kollat, & Blackwell, 1968, Ch. 8), the construct of dogmatism, as defined by Rokeach (1960) and supported in numerous subsequent investigations (cf. Vacchiano, Strauss, & Hochman, 1969), provides a sound theoretical and empirical basis for making predictions regarding innovation proneness. Consider the relationship of dogmatism to three other variables: anxiety, novelty, and creativity.

According to Rokeach (1960), dogmatism is a generalized mental rigidity factor with individuals aligned along an open- to closed-minded continuum. Highly dogmatic systems represent a cognitive-psychodynamic network of defenses against anxiety; the more persistently anxious or threatened the individual, the more likely he is to manifest a closed mind. This dogmatism-anxiety relationship has been well supported by Rokeach and several others (cf. Vacchiano et al., 1969). Consider, also, the research (Bauer, 1960; Cox, 1967) which tends to indicate that changing from one brand or product to another usually entails some perceived risk which, for several reasons, increases when one contemplates trying or switching to a new brand or product. Given that perceived risk, in anything other than mild dosages, can be anxiety inducing, high dogmatics should be less likely than low dogmatics to consider innovating.

The literature relating dogmatism to novelty is similarly suggestive. Low dogmatics score significantly higher on acceptance of change measures (Pyron, 1966a; Pyron & Lambert, 1967), react more favorably to novel music (Mikol, 1960), avant-garde art (Pyron, 1966b), and novel films (Zagona & Kelly, 1966), and function more capably in novel situations (Rokeach, Swanson, & Denny, 1960; Rokeach & Vidulich, 1960). Based on a comprehensive review of the literature, Ehrlich and Lee (1969) conclude that high dogmatics are generally more resistant to change.

It has been suggested that "innovators and early adopters tend to be more creative than later adopters [Zaltman, 1965, p. 46]." Relatedly, Rokeach (1965) predicted open-mindedness to be a necessary precondition for creativity, and Zagona and Zurcher (1965) and Jacoby (1967) have adduced supporting evidence.

Based upon both theoretical rationale and empirical evidence, it is predicted that low dogmatics will select significantly more innovations than will high dogmatics.

Last, Engel, Kollat, and Blackwell (1968) note that the most often used operational definition of innovation is "any form of a product that has recently become available in a market [p. 545]." Accordingly, totally new products (e.g., television), new brands, modifications of old brands (e.g., the annual model change on automobiles), and even a brand previously not available in a given geographical area are thus considered innovations. The present investigation utilizes this definition.

METHOD

Subjects. The *S*s were 60 18–26 year old coeds attending Purdue University during the Fall 1969 Semester. All were volunteers.

Instruments. To obtain an index of innovation proneness, 15 sets of photographs, containing 5 pictures per set, were collected for 15 different product classes. One of the five pictures in each set was of an innovation (as operationally defined above), while the remaining four pictures displayed older or more traditional varieties of product, either in terms of brand, style, or function.

For example, one product class contained pictures of four womens' dress-length coats (i.e., knee length to 5 in. above the knee) and one maxicoat (fashion innovation). Another product class, cigarettes, pictured one pack of Virginia Slims (brand innovation), Kent, Marlboro, Viceroy, and Parliament. A third product

TABLE 1

PRODUCTS AND TYPE OF INNOVATION INVESTIGATED

Product	Category	Type of innovation
1. Hair conditioner	Cosmetics	New brand
2. Lipstick	Cosmetics	New brand
3. Hair lightener	Cosmetics	Functional[a]
4. Perfume	Cosmetics	New brand
5. Blush-on	Cosmetics	Functional[a]
6. Make-up base	Cosmetics	Functional[a]
7. Men's cologne	Miscellaneous	New brand
8. Diamond ring	Miscellaneous	Fashion
9. Beer	Miscellaneous	New brand
10. Luggage	Miscellaneous	Fashion
11. Cigarettes	Miscellaneous	New brand
12. Shoes	Clothing	Fashion
13. Slack outfit	Clothing	Fashion
14. Panty-hose	Clothing	Functional[a]
15. Coat	Clothing	Fashion

[a] Functional = new form of product or method of use (e.g., foam vs. liquid make-up base).

class exhibited four regular type pantyhose and one of the lower cut hiplet variety (functional innovation).

The 15 products, classified according to category and type of innovation examined, are presented in Table 1. Across six cosmetic, four clothing, and five miscellaneous products, there were six new brands, five fashion innovations, and four functional innovations.

All five pictures in each product set were mounted in a randomized circular array on poster board so that the complete set could be viewed simultaneously. Pictures within each set were matched for size, color versus black-and-white, branded versus unbranded, photograph versus sketch, and container type (e.g., bottle vs. can).

The standard 40-item Form E of the Rokeach Dogmatism scale (Rokeach, 1960) was used to measure dogmatism.

Procedure. The *S*s who were run individually, were instructed to select one out of the five items in each set that they would buy if given the choice. The *E* continued: "Do not let the colors of the products or brands affect your decision. For example, do not choose a lipstick for the particular shade pictured. Moreover, if you do not use a set of products, buy for someone else." Each *S* received an innovation score ranging from 0 to 15, depending upon the number of times she selected the innovation contained in each of the product sets.

The 15 sets of pictures were presented in the same order, and each *S* proceeded at her own pace. The Dogmatism scale was administered upon completion—again with no time limit.

RESULTS

Based upon their dogmatism scores, *S*s were split into two groups at the median. Table 2 indicates that the between-group difference in dogmatism scores was significant at beyond the .001 level. As predicted, there was a significant overall difference ($t = 10.87$; $p < .01$, one-tailed test) in innovation scores between the two groups.

Product-by-product analysis indicated that the low dogmatics gave more innovation responses than the high dogmatics to 11 out of the 15 product sets. Four of these differences (Products 7, 9, 11, and 12 in Table 1) were significant beyond the .05 level, one-tailed test. The remaining 11 significance tests yielded probability values in excess of .20.

Table 3 provides the analysis of innovative responses according to both type of product and type of innovation. Low dogmatics were considerably more likely ($p < .001$) than high dogmatics to select new brands of products, especially in the miscellaneous category ($p < .05$).

DISCUSSION

As predicted, low dogmatic individuals tend to make significantly more innovative selections than do high dogmatic individuals. However, the relationship, while significant, does not appear to be exceptionally strong. Only 10% of the variance is accounted for by the correlation between dogmatism and innovation proneness ($r = -.32$). Moreover, the mean low dogmatic innovation score of 3.27 is only .27 above, and the high dogmatic mean only .83 below, the 3.00 value to be expected on a chance basis (1/5 innovations × 15 product sets).

TABLE 2

SIGNIFICANCE OF THE DIFFERENCE IN MEAN INNOVATION SCORES OF HIGH AND LOW DOGMATICS

Innovation scores	Low dogmatics[a]	High dogmatics[a]	t
Dogmatism			
$\bar{X}$	118.40	161.13	10.87**
SD	14.53	15.90	
Difference			
$\bar{X}$	3.27	2.17	2.53*
SD	2.03	1.23	

Note.—Dogmatism × Innovation Index: $r = -.32$, $p < .01$.

[a] $n = 30$.

* $p < .01$ (given heterogeneity of variance and using $df = 29$ [cf. Edwards, 1960, p. 108], one-tailed test).

** $p < .001$ (one-tailed test).

From an applied viewpoint, however, the notion of practical significance is more important than statistical significance or even the size of the relationship. To the extent that marketers and advertisers can capitalize on knowledge regarding the relationship between dogmatism and innovation proneness, and can thereby increase the likelihood of gaining that all-important toehold, the relationship is important. How, then, might this information be applied?

It has been suggested that when introducing new products, "advertisements should emphasize the prestige of owning the item—to appeal specifically to the early adopters [Zaltman, 1965, p. 52]." Given that low dogmatics are more innovative and tend to adopt earlier than high dogmatics, then quite the opposite prediction should be made. In describing the essential differences between high and low dogmatics, Rokeach (1960) states that low dogmatics "act on relevant information from the outside of its own intrinsic merits, unencumbered by irrelevant factors in the situation arising from within the person or from the outside [p. 57]." Irrational egomotives and the need for self-aggrandizement are cited by Rokeach as examples of irrelevant internal factors, while pressures arising from peers, reference groups, authority figures, social, institutional, and cultural norms are cited as examples of irrelevant external factors. High dogmatics prefer to rely heavily on the pronouncements of authorities and tend to accept information primarily on the basis of who said it (i.e., the message source), while low dogmatics tend to act based upon their independent evaluation of the information itself (i.e., message content). Substantial support for these relationships are to be found in the social-psychological literature (cf. Ehrlich & Lee, 1969; Powell, 1962; Rokeach, 1960; Vacchiano et al., 1969).

Translated into applied terms, this suggests that the marketer-advertiser introducing a new product would be wise to have his advertisements and point-of-purchase displays focus on the product's attributes rather than emphasize irrelevant ego or social factors. As a hypothetical example, advertisers of a cigarette with a new type of filter (ergo, a new product) would do well to stress and describe the filter rather than emphasize people enjoying the cigarette in social situations. Such differential reactivity to prestige versus factual appeals could be tested using an approach similar to Kassarjian (1965).

Last, it would appear that a multiple-indicator approach, especially when coupled with theoretically based a priori hypotheses, has great promise for the study of innovators. Had a single indicant criterion (i.e., studying the characteristics of innovators based on their responses to only one innovation) been utilized, the predicted significant relationship would have been manifested in only 4 out of 15 instances, and the likelihood of concluding that no relationship existed thereby increased.

In conclusion, low dogmatic individuals were found to be significantly more likely to be innovators than high dogmatic individuals. To the extent that such innovators are also more likely to prefer objective information, this suggests that advertisers seeking to establish a market position for an innovation ought to focus on presenting factually oriented information rather than emphasizing irrelevancies.

TABLE 3

FREQUENCY OF INNOVATIVE RESPONSES BY TYPE OF PRODUCT AND TYPE OF INNOVATION

Type	Low dogmatics	High dogmatics	χ^2
Product			
Cosmetics[a]	31	23	1.19
Clothing[b]	17	11	1.29
Miscellaneous[c]	50	31	4.46*
Innovation			
New brand[a]	48	20	11.53**
Fashion[c]	29	28	.69
Function[b]	21	17	.42

[a] $N = 6$.
[b] $N = 4$.
[c] $N = 5$.
* $p < .05$ (one-tailed test).
** $p < .001$ (one-tailed test).

REFERENCES

BAUER, R. A. Consumer behavior as risk-taking. *Proceedings of the American Marketing Association*, 1960, 389–398.

BOHLEN, J. M., COUGHENOUR, C. M., LIONBERGER, H. F., MOE, E. O., & ROGERS, E. M. Adopters of new farm ideas: Characteristics and communications behavior. In H. H. Kassarjian & T. S. Robertson (Eds.), *Perspectives in consumer behavior*. Glenview, Illinois: Scott, Foresman, 1968.

COLEMAN, R. P. The significance of social stratification in selling. *Proceedings of the American Marketing Association*, 1960, 171–184.

COX, D. F. (Ed.) *Risk taking and information handling in consumer behavior.* Boston: Graduate School of Business Administration, 1967.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Holt, Rinehart & Winston, 1960.

EHRLICH, H. J., & LEE, D. Dogmatism, learning, and resistance to change: A review and a new paradigm. *Psychological Bulletin*, 1969, **71**, 249–260.

ENGEL, J. F., KEGERREIS, R. J., & BLACKWELL, R. D. Word of mouth communication by the innovator. *Journal of Marketing*, 1969, **33**, 15–19.

ENGEL, J. F., KOLLAT, D. T., & BLACKWELL, R. D. *Consumer behavior.* New York: Holt, Rinehart & Winston, 1968.

FRANK, R. E., MASSY, W. F., & MORRISON, D. G. The determinants of innovative behavior with respect to a branded, frequently purchased food product. *Proceedings of the American Marketing Association*, 1964, 312–323.

JACOBY, J. Open-mindedness and creativity. *Psychological Reports*, 1967, **20**, 822.

KASSARJIAN, H. H. Social character and differential preference for mass communications. *Journal of Marketing Research*, 1965, **2**, 146–153.

KING, C. W. Fashion adoption: A rebuttal to the "trickle down" theory. *Proceedings of the American Marketing Association*, 1963, 108–125.

MIKOL, B. The enjoyment of new musical systems. In M. Rokeach (Ed.), *The open and closed mind.* New York: Basic Books, 1960.

POWELL, F. A. Open- and closed-mindedness and the ability to differentiate source and message. *Journal of Abnormal and Social Psychology*, 1962, **65**, 61–64.

PYRON, B. A factor analytic study of simplicity-complexity of social ordering. *Perceptual and Motor Skills*, 1966, **22**, 259–272. (a)

PYRON, B. Rejection of avant-garde art and the need for simple order. *Journal of Psychology*, 1966, **63**, 159–178. (b)

PYRON, B., & LAMBERT, P. The generality of simplicity-complexity of social perception in a high school population. *Journal of Psychology*, 1967, **66**, 265–273.

ROBERTSON, T. S. Social factors in innovation behavior. In H. H. Kassarjian & T. S. Robertson (Eds.), *Perspectives in consumer behavior.* Glenview, Illinois: Scott, Foresman, 1968.

ROGERS, E. M. *Diffusion of innovations.* Glencoe, Illinois: The Free Press, 1962.

ROGERS, E. M., & STANFIELD, J. D. Adoption and diffusion of new products: Emerging generalizations and hypotheses. In F. M. Bass, C. W. King, & E. Pessemier (Eds.), *Applications of the sciences in marketing management.* New York: Wiley, 1968.

ROKEACH, M. *The open and closed mind.* New York: Basic Books, 1960.

ROKEACH, M. In pursuit of the creative process. In G. A. Steiner (Ed.), *The creative organization.* Chicago: University of Chicago Press, 1965.

ROKEACH, M., SWANSON, T. S., & DENNY, M. R. The role of past experience: A comparison between chess players and non-chess players. In M. Rokeach (Ed.), *The open and closed mind.* New York: Basic Books, 1960.

ROKEACH, M., & VIDULICH, R. N. The formation of new belief systems: The roles of memory and the capacity to entertain. In M. Rokeach (Ed.), *The open and closed mind.* New York: Basic Books, 1960.

VACCHIANO, R. B., STRAUSS, P. S., & HOCHMAN, L. The open and closed mind: A review of dogmatism. *Psychological Bulletin*, 1969, **71**, 261–273.

ZAGONA, S. V., & KELLY, M. A. The resistance of the closed mind to a novel and complex audio-visual experience. *Journal of Social Psychology*, 1966, **70**, 123–131.

ZAGONA, S. V., & ZURCHER, L. A. Participation, interaction and role behavior in groups selected from the extremes of the open-closed cognitive continuum. *Journal of Psychology*, 1965, **60**, 213–219.

ZALTMAN, G. *Marketing: Contributions from the behavioral sciences.* New York: Harcourt, Brace & World, 1965.

(Received March 2, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 4, 389–392

COMPARATIVE STUDY OF ARTISTS AND NONARTISTS ON THE MAITLAND GRAVES DESIGN JUDGMENT TEST[1]

H. J. EYSENCK AND M. CASTLE[2]

Institute of Psychiatry, University of London

The Maitland Graves Design Judgment Test was administered to 229 male and 236 female controls (nonart students), and 519 male artists and 478 female artists (art students). Only very slight differences were observed between the experimental and control groups. Following the lines of a previous factor analysis of the test, scores were derived for three subsets of designs (symmetrical, three-dimensional, complex); these scores were negligible correlated with each other. Conspicuous differences between artists and controls were only found for the first of these sets; it was also found that scores on this set were more homogeneous for artists than for controls. The use of overall scores on a heterogeneous test is criticized, and the development of a more homogeneous test based on symmetrical designs only is suggested. An explanation is suggested for the differences in results reported by Graves and the present authors.

The Maitland Graves Design Judgment Test (1947) is based on certain theories of artistic creation and appreciation (Graves, 1941). It contains 90 pairs or triads of designs one of which has to be chosen by *S* as aesthetically superior to the other(s). One particular choice in each case is considered "right" (and scored + accordingly) on the basis of agreement with the author's theories and the agreement of art teachers on the superiority of that particular design. Graves (1948) reports that art students achieve higher scores on the test than nonart students; mean scores for college students are 76 and 48, respectively, for high school students, 56 and 38, respectively. These values, taken from Graves (1948, p. 3, Table 1) are typical of others given in Graves (Tables 2 & 3, p. 4). The author concludes that "the test's ability to differentiate the art groups from the non-art groups is unmistakably clear [p. 2]."

The theories advanced by Graves are of considerable psychological interest, and if his test did in fact succeed in differentiating so clearly between artists and nonartists, this would have important implications for a theory of aesthetics quite apart from the practical value of the test as an instrument of selection.

Impressive as these figures are, there are nevertheless certain doubts that require to be set at rest before one can accept the test as a measure of aesthetic aptitude. In the first place, the data presented are restricted to American students at one point of time; the possibility cannot be ruled out that artists in other countries, or even in the United States at other periods of time may show different preferences to those endorsed by Graves's (1948) experts. The alleged objectivity of aesthetic criteria has not been demonstrated sufficiently to allow us to disregard such a possibility, and accordingly, the present study has repeated Graves's attempt with a sample of art and nonart students differing from his by being citizens of the United Kingdom, and by being tested some 25 yr. later. The authors have also taken care to keep separate male and female students, as the possibility cannot be gainsaid that sex differences may interfere with the establishment of clear-cut art versus nonart student differences; little is known about this.

In the second place, Graves (1948) has not established certain important psychometric properties of his scale. This scale is considered to be a measure of a single aesthetic property, but no evidence is offered of this proposition. Graves claims as one of his criteria for retention of items "greater preference for the design by those who achieved high scores on the entire test than by those receiving low scores (internal consistency) [p. 2]," but by itself this criterion is of doubtful value, and in any case no details of its application have been published. Eysenck (1967) has published details of a factor analytic study of the 90 items of the test, and has shown that several different factors appear to determine choice of design, rather than the one single factor suggested by Graves. Three factors appeared to be particularly worthy of study in this connection. Factor 1 (symmetry) is best represented in Items 8, 21, 28, 29, 47, 48, 50, 51, 57, 59, 61, 69, and 72, all of which had loadings of above .5 on this factor. These items are all concerned with pairs of designs in which one shows excessive symmetry, while the other (preferred) design departs from strict symmetry. Factor 2 (three-dimensional) consists of Items 80–90, all of which are three-dimensional, in contrast to the other items which are two-dimensional. Factor 3 (complexity) consists of Items 1, 12, 26, 40, 55, and 74; these are all characterized by being very complex drawings, in contrast with the essentially simple drawings which make up Factor 1. Scores were obtained for all our groups on these three factors, using Graves's key, as well as on the total scale.

---

[1] The authors are indebted to the Social Science Research Council for a grant which made this study possible.

This article was accepted for publication by the outgoing Editor, Dr. Kenneth E. Clark.

[2] Requests for reprints should be sent to H. J. Eysenck, Department of Psychology, Institute of Psychiatry, University of London, S. E. 5, England.

Four groups were tested in all: male art students ($N = 519$), female art students ($N = 478$), male nonart students ($N = 229$), and female nonart students ($N = 236$). The Ss were divided into artists and controls on the basis of whether or not they had received formal training in the visual arts. To make more precise our use of terms, we will list the types of courses that those groups under the term "art students" had attended, or were attending. These courses included graphic design, fashion design, fine arts (painting and sculpture), architectural studies, typographic design, textile design, cinematography, photography, industrial design, theater design, interior design, stained glass and ceramic design, silver and jewelry design; included were also some post-graduate students and professional artists. By contrast, the controls included students of surveying, law, commerce, accounting, hairdressing, electrical engineering, mechanical engineering, secretarial students, hotel work and catering students, student linguists, gas fitting apprentices, and day release (retail trade) students. Ages of both groups ranged from 16 to 25 yr., with only a few above the latter age; the great majority were around 20 yr. old. The groups differed with respect to their preference judgments on a test of polygonal forms (Eysenck & Castle, 1970), with the artists preferring the simpler forms. The artists were also found to have higher scores on the N scale of the Eysenck Personality Inventory, but not on the E scale.

TABLE 1

SCORES OF ARTISTS AND CONTROLS ON TOTAL TEST AND THREE SUBTESTS

Test or subtest	Control		Artist	
	Male[a]	Female[b]	Male[c]	Female[d]
S_1 (Total)	52	56	58	58
S_2 (Symmetry)	7.3	8.1	9.4	9.3
S_3 (Three-dimensional)	6.5	7.4	6.6	7.1
S_4 (Complexity)	3.6	2.9	3.8	3.6

[a] $n = 229$.
[b] $n = 236$.
[c] $n = 519$.
[d] $n = 478$.

Presentation of the material was by way of slides shown to medium-sized groups of students. Close supervision ensured that no cheating or talking took place. Seating was arranged in such a way that little angular distortion occurred. It does not seem likely that this variation from the traditional method of administration of the test would have given rise to differences in scoring, although of course it is impossible to be certain of this.

Results of the experiment are shown in Table 1. It is clear that total scores give minimal differentiation between artists and controls, with the former scoring lower than Graves's artists, and the latter higher than his controls. Analysis of variance shows the data to be significant beyond the .001 level, but this is not unexpected in view of the large number of Ss. Male artists are significantly superior to male controls ($p < .001$), but female artists are not significantly different from female controls. Female artists are significantly superior to male controls ($p < .001$), but male artists are not significantly different from female controls. All differences are rather slight, and cannot compare with those reported by Graves. Our mean difference, weighing the samples equally, is 4 points, while that reported by Graves was 28 points. Clearly, our study has essentially failed to replicate Graves's original findings.

Scores on the symmetry factor show a somewhat different picture. Here artists are clearly superior to controls, for both sexes; in each case $p < .001$. Female controls are superior to male controls ($p < .01$), but not female artists to male artists. Scores on the three-dimensional factor show an overall level of significance of $p < .001$, but this is entirely due to the superiority of the female *S*s over the male *S*s ($p < .001$ for controls, and $p < .01$ for artists); male artists do not differ from male controls, and female controls are actually superior ($p < .05$) to female artists. Scores on the complex designs again show overall differences at the $p < .001$ level. Female artists are better than female controls ($p < .001$), but male controls are not differentiated from male artists. Male controls are superior to female controls ($p < .001$), but male artists are not superior to female artists. These data demonstrate that it is essential to treat men and women differently in the analysis of the data, and also that the factors isolated in the Graves (1948) Design Judgment Test must be analyzed separately, as otherwise much important information is lost.

TABLE 2

INTERCORRELATIONS OF MALE CONTROLS (MC), FEMALE CONTROLS (FC), MALE ARTISTS (MA), AND FEMALE ARTISTS (FA) FOR SUBSETS

Subset	S_2	S_3	S_4
S_1			
MC	.85	.17	.25
FC	.88	.16	.38
MA	.82	.35	.35
FA	.81	.27	.40
S_2			
MC		−.04	.03
FC		−.02	.20
MA		.14	.16
FA		.05	.26
S_3			
MC			−.00
FC			.05
MA			.07
FA			.03

The original factor analysis had been done on a group of nonartists, and it is possible that for artists the factors there demonstrated might not be applicable. If this were so, then the correlations between subsets (S_2, S_3 and S_4) would be expected to be substantially larger than zero for the artists, and significantly different from the control group. The observed correlations between these sets are given in Table 2, together with the correlations of each set and S_1, the total score. The total score, of course, incorporates the subscores, so that these correlations are artificially inflated. However, as the subsets are roughly of equal size, comparisons between them are still meaningful and possible. It will be noted that for all four groups the intercorrelations between S_2, S_3, and S_4 are very low, and even on occasion negative; this substantiates the validity of the original factor analysis. There is no evidence here that artists have higher correlations than controls. Correlations of S_1 with the other scales shows that S_2 is very highly correlated with the total scale, but that the correlations of S_3 and S_4 with S_1 are quite low and hardly above the level of the artificial correlation produced by the fact that the items involved in these scales also appear in S_1. We may conclude that S_3 and S_4 not only fail to differentiate properly between artists and controls, but that they also fail to correlate with S_2, which does so discriminate. Clearly, whatever validity the total scale may possess derives essentially from the items in S_2; items making up S_3 and S_4 could with advantage be jettisoned as they contribute nothing but error and unreliability. The short 13-item S_2 scale discriminates better between artists and controls than does the whole 90-item scale including these 13 items.

It seems likely that the artist group is more homogeneous than the control group for aesthetic ability, and a look at the variances of the various scores used may be useful. S_3 and S_4 do not show any significant differences; S_1 does so at a $p < .05$ value. Most striking, however, is the difference in variances for S_2, which is significant at the $p < .001$ level. The actual variances are: 3.75^2 (male controls), 3.61^2 (female controls), 2.84^2 (male artists), 2.87^2 (female artists). It is clear that on those subsets of items which discriminate best between artists and controls, the artists have lower variances than the controls. This is very much what one might have expected. Some psychologists would argue that such differences in variance make inapplicable the *F* test and *t* test used, but mathematicians would not uphold such nvestigating the possible interaction between ratio of value and exposure duration.

METHOD

Subjects

The *Ss* were 80 male volunteers from Syracuse University. The *Ss* competed for monetary awards of \$10.00, \$5.00, \$3.00, and \$2.00 that were given to the four best performers, as determined by the number of points accumulated, within each group of eight *Ss*.

Apparatus

The stimulus symbols were typewritten capital letters from the alphabet. To allow for a comparison between high- and low-value stimuli, the alphabet was split into the subsets of A–J and Q–Z. The letters K–P were never presented in order to provide a clear distinction between the subsets.

The stimulus displays consisted of 2×2 negative slide projections with either 4, 8, or 12 white letters on a black background. Each level of load, that is, the number of letters per slide, was replicated 15 times with a different random sample of the alphabet for a total of 45 test slides. Each letter appeared an equal number of times during the 45 trials, and all slides contained an equal number of letters from each stimulus subset. Following selection, the letters were randomly assigned to a location in a 10×10 rectangular matrix with the restriction that half the letters selected from each subset be placed in the center 36 positions of the matrix and the other half placed in the 64 peripheral positions. The letters were then typed in appropriate locations on sheets of white, unlined paper that were photographed for slides.

A Kodak Carousel slide projector with an externally mounted shutter presented the slides, while automatic programming equipment controlled the exposure durations and interslide intervals. The projected size of the matrix area of slides was 43×60 in., while the projected size of the letters was 1 in. The *Ss* sat in student-type arm chairs arranged in two rows of four each with the average distance from *Ss*' eyes to the screen being 12 ft. from Row 1 and $15\frac{1}{2}$ ft. from Row 2.

Procedure

Eight *Ss* were assigned to each of 10 groups derived from the combination of five stimulus exposure durations and two value ratios. The exposure durations used were .5, .75, 1.0, 2.0, and 3.0 sec. per slide. The ratios between high- and low-value letters were 2:1 and 16:1 with one point always given for each correct report of a letter in the low-value subset and either 2 or 16 points given for each of the correctly reported letters from the high-value subset. Four *Ss* within each group of eight performed with the subset from the first half of the alphabet (A–J) as the high-value categories, while the other four had Q–Z as the high-value categories. All *Ss* viewed the identical slides and thus performed under all levels of load (4, 8, and 12 letters per slide).

The *Ss* served for approximately 45 min. with the first 15 min. reserved for instruction and practice. The *Ss* were informed that letters could appear at all places on the screen and that points would be received for each letter reported correctly in their answer booklet. All *Ss* were further instructed that the number of points received for any particular letter was equal to the value assigned to the population in which it was contained, and that aside from the possibility of being dropped for random guessing, there would be no penalty or loss for incorrect answers. A card, permanently in view, on *Ss*' answer booklets informed them of the ratio of value and of the designated values for each population subset.

Nine practice trials were given with the appropriate exposure duration and intertrial interval to insure that each *S* understood the task. On four of the trials, the slide was reexposed to allow *S* to tally the number of points he accumulated. Following practice, all *Ss* viewed the 45 slides in the same randomly presented order, were given a short rest, and then given the same 45 slides in a new order. The *Ss* were allowed 10 sec. for each report. This was followed by a .5-sec. warning buzzer and a 3-sec. interval to allow *Ss* to adjust the answer booklet for the next presentation. No feedback was provided during the testing period.

RESULTS

The data were analyzed using a number of different response measures to assess *Ss*' strategy and ability for processing information of different value levels. Although there are a considerable number of possible response measures, those selected for analyses and presentation were those which appeared to produce inconsistent results in the previous literature. Although blocks of trials (Trials 1–45 vs. Trials 46–90) were analyzed and were a significant variable with some of the measures, these findings are not presented, since they do not alter the general findings and conclusions of the major variables of interest.

Initial Responses

These data were first analyzed for the percentage of times that a high-value letter was the first letter recorded by an *S* regardless of the accuracy of the report (Christ, 1969; Taub, 1965). Each percentage was based on a maximum of 15 possible responses at each level of load for each block of trials. The results of an analysis of variance indicated that the percentage of high-value initial reports increased directly with exposure duration ($F = 15.70$, $df = 4/70$, $p < .01$), ratio ($F = 14.57$, $df = 1/70$, $p < .01$) and their combination ($F = 2.61$, $df = 4/70$, $p < .05$). Figure 1, a plot of the

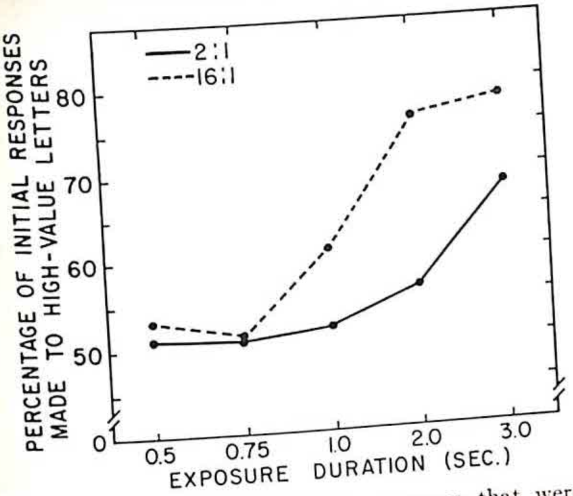


Fig. 1. Percentage of initial responses that were high-value letters as a function of exposure duration and ratio.

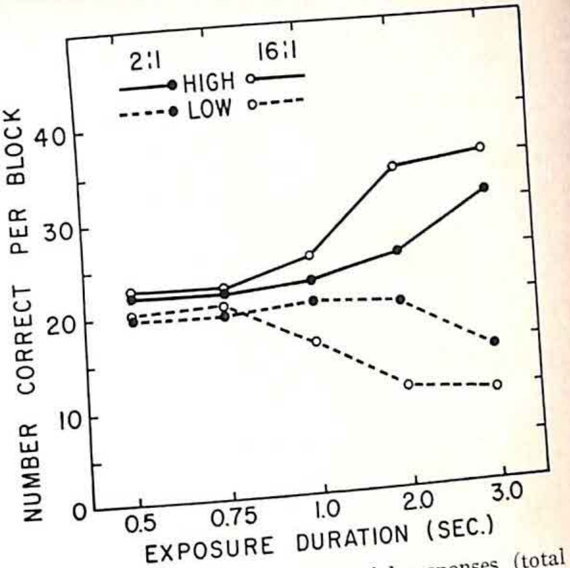


Fig. 2. Number of correct initial responses (total maximum = 45) per block of trials as a function of value, ratio, and exposure.

Ratio × Exposure interaction, shows that at .5 and .75 sec. durations *S*s made initial responses of high-value stimuli only about 50% of the time with both ratios, and that although the proportions of these responses increased directly with exposure, increases were larger for those *S*s performing with a 16:1 ratio. The significant effects involving load indicated that the percentage of high-value initial responses increased with the number of letters on a slide ($F = 25.95$ $df = 2/140$, $p < .01$) and that the size of the differences among loads varied directly with exposure duration ($F = 6.08$, $df = 8/140$, $p < .01$). The means underlying the main effect of load were 52.88% at 4 letters, 62.59% at 8 letters, and 63.67% at 12 letters per slide, while the respective means were 51.34%, 51.34%, and 53.19% at .5 sec. and 56.47%, 80.06%, and 82.88% at 3.0 sec. No other interactions were significant with this performance measure.

As a check on the findings for exposure and ratio, the data were also scored and analyzed for the number of correct initial reports of high- and low-value letters (Hearns & Moss, 1968). In general, these findings were consistent with those of the previous analysis and indicated that the following effects were significant: exposure ($F = 3.83$, $df = 4/70$, $p < .05$), value ($F = 84.15$, $df = 1/70$, $p < .01$), Value × Exposure ($F = 14.29$, $df = 4/70$, $p < .01$), Value × Ratio ($F = 11.95$, $df = 1/70$, $p < .01$), and Value × Exposure × Ratio ($F = 2.86$, $df = 4/70$, $p < .05$). Figure 2, a plot of the latter interaction, indicates that the shape of the curves for high-value responses are identical to those presented in Figure 1, while those for low-value responses are represented by the mirror images.

Total Reports

Table 1 presents a summary of the results of the analyses of variance that were run for each of the measures of total performance. Analysis

TABLE 1

Summary of Degrees of Freedom, *F* Values and *p* Values for Each of the Analyses of Total Reports

Source	*df*	% correct transmissions	No. responses	Accuracy of reports
Exposure duration (E)	4	63.87**	35.83**	5.74**
Ratio (R)	1	.75	5.67*	1.67
E × R	4	.88	1.80	.75
Ss / E × R	70			
Value (V)	1	105.20**	131.74**	4.86*
V × E	4	23.29**	22.81**	5.62**
V × R	1	14.92**	14.68**	.18
V × E × R	4	3.43*	3.08*	1.54
Ss × V / E × R	70			
Load (L)	2	1779.81**	331.00**	19.89**
L × E	8	24.16**	14.71**	20.22**
L × R	2	2.30	5.15**	1.15
L × E × R	8	.58	1.65	1.07
Ss × L / E × R	140			
L × V	2	52.78**	62.27**	3.50*
L × V × E	8	16.19**	16.48**	.88
L × V × R	2	5.55**	6.49**	4.28*
L × V × E × R	8	2.52*	2.60*	.81
Ss × L × V / E × R	140			

* $p < .05$.
** $p < .01$.

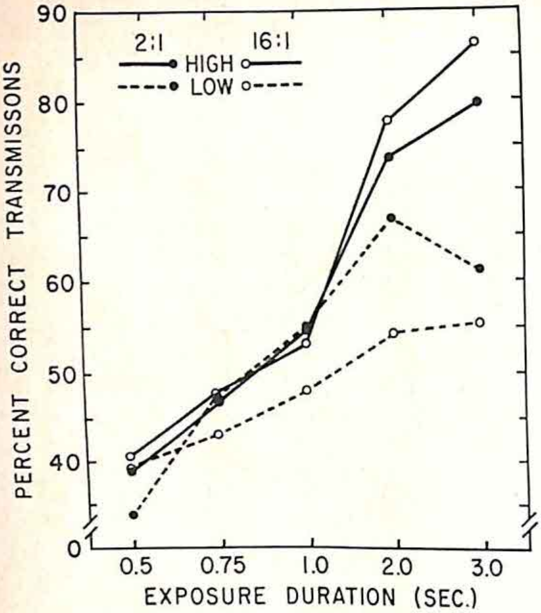


FIG. 3. Percent-correct transmission as a function of value, ratio, and exposure.

of percent-correct transmissions of high- and low-value stimuli, defined as the number correct divided by the number of stimuli presented (Christ, 1969; Taub, 1965) indicated that more of the high-value than low-value stimuli were transmitted correctly and that the differences between high- and low-value transmissions increased with both exposure duration and ratio of value. Although the main effect of ratio was not significant, the data underlying the significant Value × Ratio interaction suggested that as ratio was increased, the major trend was toward a decrement in transmissions of low-value stimuli. However, Figure 3, a summary of the Value × Exposure × Ratio interaction, shows that this trend was not present at all exposure durations. This latter interaction indicated that while the 16:1 ratio only led to decreases in low-value transmissions when compared to the 2:1 ratio at the .75- and 1.0-sec. exposure durations, there was an increase in high-value transmissions as well as a decrease in low-value transmissions at the longer durations of 2.0 and 3.0 sec.

Table 1 further shows that a number of effects involving load were significant with this performance measure. The means underlying the main effect for load indicated that 79.4%, 49.5%, and 36.3% of the stimuli were transmitted correctly with 4, 8, and 12 letters per slide, respectively. The data underlying the significant interaction of Load × Exposure indicated that the differences with load increased with exposure duration, while interactions involving Load × Value and/or Ratio were mainly due to the finding that these variables affected transmission only with loads of 8 and 12 letters per slide. Table 2 presents the scores for the Load × Value × Ratio interaction and further illustrates the data underlying the Value × Ratio interaction.

Analysis of the number of responses of high- and low-value stimuli regardless of accuracy (Christ, 1969; Hearns & Moss, 1968) revealed that the trends underlying the findings for exposure, value, and ratio were identical in

TABLE 2

PERCENT-CORRECT TRANSMISSION AND MEAN NUMBER OF RESPONSES PER TRIAL AS A FUNCTION OF LOAD, VALUE, AND RATIO

Measure	Value	Letters per slide					
		4		8		12	
		Ratio					
		2:1	16:1	2:1	16:1	2:1	16:1
% transmission	High	79.16	80.44	55.49	57.98	41.58	44.66
	Low	79.16	78.84	45.59	38.80	33.41	25.55
$\bar{X}$ no. responses per trial	High	1.80	1.78	2.55	2.56	2.82	2.90
	Low	1.78	1.68	2.11	1.76	2.22	1.69

shape to those presented in Figure 3. Similarly, for the Value × Ratio interaction, the 2:1 and 16:1 ratio groups made 2.38 and 2.41 high-value responses and 2.03 and 1.71 low-value responses, respectively. However, unlike with percent-correct transmissions, there was a small but significant difference due to ratio with less responses per trial being made at 16:1 (4.13) than at 2:1 (4.43). As in the previous analysis, there were a number of significant effects due to load. The data underlying the main effect indicates that the number of responses increased directly with load with 3.52, 4.49, and 4.81 responses made with 4, 8, and 12 letters per trial, respectively. Further, although the direction of effects are opposite to those with percent-correct transmission, the underlying differences are the same. That is, the difference due to load increased with exposure duration, while value and ratio had no effect on the number of responses with four letters but produced differences at both 8 and 12 letters. An example of the data underlying these effects are presented in Table 2.

Analysis of the accuracy of reports, defined as the number of correct responses divided by the total number of responses (Hearns & Moss, 1968), suggested that the variables had less of a consistent effect on this measure than on the other measures of performance. The significant effects due to exposure, value, and their interaction indicated that accuracy increased directly with exposure duration (83.14%, 86.30%, 86.57%, 92.76%, and 95.18%) and that although accuracy was slightly higher for the low-value letters (88.23% for high value, and 89.35% for low value), this trend was only present for 1 sec. and shorter durations with a slight reversal in favor of high-value letters at 2.0 and 3.0 sec. Ratio had little consistent effect and did not significantly affect accuracy either alone or in combination with exposure or value. The interaction of Ratio × Value × Load was the only significant effect with ratio and appeared to be due to a slight reversal in relative accuracy for high-value (91.4%) and low-value (90.9%) stimuli for the 16:1 ratio condition with 12 letters per trial. Load significantly affected accuracy, although the means of 88.95% at 4 letters, 87.00% at 8 letters, and 90.42% at 12 letters indicated that the trend across load was not consistent. Finally, the Load × Exposure interaction suggested that accuracy varied directly with load at .5, .75, and 1.0 sec. and inversely with load at 2.0 and 3.0 sec.

DISCUSSION

In general, the present results are consistent with suggestions that differential ratio of value has an effect on performance in a visual search task (Christ, 1969; Taub, 1965; Tolkmitt & Christ, 1970). The data underlying the major interactions with initial responses, mean number of responses per trial, and percent-correct transmissions indicate that the differences in performance as a function of ratio increased as exposure duration increased. The shape of the interactions suggest that with short exposure durations the differences in percentage of stimuli correctly transmitted and number of responses per trial were, to a large extent, due to an inhibition of low-value responses as ratio was increased, a finding which is consistent with the trends observed by Christ (1969) and Taub (1965). However, although the trends are similar, the present analyses as well as those of Christ (1969) and Taub (1965) suggest that the main effect of ratio may or may not be significant depending on the relative size of the increase in responses of high-value stimuli compared to the inhibition of low-value responses as ratio is increased. The present results extend these findings by suggesting that the amount of increase in high-value responding with increases in ratio may be directly dependent on duration of exposure. This latter result is consistent with the findings recently reported in a comparison of an equal value (1:1) with a differential value (4:1) condition at .5-sec. and 2.5-sec. exposures (Tolkmitt & Christ, 1970).

The current findings of interactions for exposure, ratio, and value with initial responses, mean number of responses per trial, and percent-correct transmissions suggest that a large part of the inconsistencies among studies of differential ratio of value with a visual search task may be related to the exposure durations that were used. Figures 1–3 clearly demonstrate that differences due to ratio are small, if present at all, at .5 and .75 sec., which would agree with the reported negative findings of Hearns & Moss (1968). Further, although Christ (1969) did not find any differences with

INFORMATION FOR CONTRIBUTORS

Style. A professional article should possess certain characteristics: (*a*) conciseness and an apparent respect for reader time; (*b*) unambiguous and simple vocabulary with technical and erudite words used only when simpler ones would obviously be inadequate; (*c*) conformity to accepted technical style in tables, terminology, and references; (*d*) conclusions that are clearly related to the evidence presented. The reader should be led, step by step, from a statement of problem or purpose, through analysis of evidence, to conclusions and implications. Authors are encouraged to consult the little book entitled *The Elements of Style* by W. Strunk, Jr. and E. B. White (New York, Macmillan, 1962).

Criteria for evaluation. Manuscripts will be evaluated on the basis of several criteria, including: (*a*) significance in contributing new knowledge to the field, (*b*) technical adequacy, (*c*) appropriateness for the *Journal of Applied Psychology*, and (*d*) clarity of presentation.

Format. Manuscripts must be prepared in the format described in the *Publication Manual of the American Psychological Association* (1967 Revision), obtainable for $1.50 from the American Psychological Association, 1200 17th Street, N.W., Washington, D. C. 20036. Articles not prepared in this manner cannot be reviewed. Special attention should be given to the section on typing the manuscript (p. 48) and to the sections on tables, figures, and references. Note that *all copy must be double spaced*, including references, title, figure captions, etc. The senior author's last name should appear on each page in the upper left-hand corner. In reference lists, give journal titles in full; do not abbreviate.

Copies. All manuscripts must be submitted in triplicate. One copy should be the original typed copy, the other two, clear carbon copies or photo reproductions, only. Authors should check the final typing carefully and retain a copy of the manuscript as a precaution against loss in the mail.

Length. Before preparing a manuscript, the author should check several recent issues to get an idea of the approximate length of regular articles published in the *Journal of Applied Psychology.* (One printed page equals roughly three double-spaced typewritten manuscript pages.) A few longer articles of special significance may be printed from time to time as monographs. Occasionally the Journal will have a section of "Short Notes" featuring brief reports on studies which make some methodological contribution or constitute an important replication.

Abstracts. Each copy of the manuscript must be accompanied by an abstract of 100–120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts.* Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or they may be obtained from the Editor or from the APA Central Office.

Figures. The original drawing of a figure, or an 8 × 10-in. glossy print of the drawing, is needed for publication and must be submitted with the original typed copy of the manuscript. Duplicate copies of the figure may be photographic or pencil-drawn copies. Figures must be hand-lettered; typewritten lettering is not acceptable.

Early publication. Regular articles are published as nearly as possible in the order in which they were received. Although everything possible is done to keep publication lag to a minimum, at present there is a 10–12-mo. period between the time of receipt and publication of an article. The author of an accepted manuscript may secure early publication, usually in the first issue to go to press after its acceptance, by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication will be sent with the letter of acceptance.

Reprints. No gratis reprints are supplied. Reprints may be ordered from the printer prior to publication.

National Auxiliary Publications Service (NAPS). To save journal space, while at the same time making all necessary materials available to those with a special interest in an article, supplementary materials (such as extensive tables or copies of instruments) may be deposited with the NAPS. Copies of the material to be deposited should be submitted to the Editor with the manuscript and its availability indicated with the following footnote in the article:

> [3] *A five-page table giving data for each of the subgroups* (or other described material) has been deposited with the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

Journal of Applied Psychology
1971, Vol. 55, No. 5, 399–405

RELATIONSHIP BETWEEN AGE AND RISK TAKING AMONG MANAGERS

VICTOR H. VROOM[1] AND BERND PAHL

Graduate School of Industrial Administration, Carnegie-Mellon University

A short version of Kogan and Wallach's choice-dilemma test was administered to 1,484 managers from over 200 companies. The results show a significant negative relationship between age and both risk taking and the value placed upon risk. The relationships are small in size when individual scores are correlated with age, but increase markedly when mean scores are used. Furthermore, the relationships are relatively stable across items and across companies. Developmental and sociocultural mechanisms that might underlie these findings are discussed.

The concept of risk taking has played a central role in psychological research during the last decade. A large amount of research has been conducted in an effort to determine the situational determinants of risk taking, including the mechanisms involved in the frequently reported finding that group decisions tend to be riskier than those of individuals.

Somewhat less attention has been paid to individual differences in risk taking, including its relationship to age. Age is a frequently used variable in psychological research and its relationship to a large number of diverse behaviors has been studied (Wohlwill, 1970). But little is known about age and risk taking, particularly among mature adults. In one of the few studies dealing with this question, Wallach and Kogan (1961) compared risk-taking behavior on their choice-dilemma tasks of college age and elderly men and women. The older *S*s, both males and females, were found to be significantly more conservative than the college students. They attribute the observed differences to the financial anxieties that affect the aged in our society.

In a later review article (Kogan & Wallach, 1967) the authors comment on the need for further exploration of the age–risk-taking relationship using less extreme age groups.

> Although risk-taking comparisons of college students and elderly individuals yield results of some interest, it would be quite valuable to collect data at age intervals over the entire adult life span. There is a striking dearth of information about adults who are beyond college and not yet eligible for gerontological research. There are the people who are least accessible, of course, because of occupational and family commitments. If we could collect systematic data on subjects in the thirty to sixty age range, we would obviously be able to specify a reasonably accurate age function for risk-taking. Is it a progressively declining smooth curve, or are there peaks and valleys [p. 168]?

The relationship between age and risk taking is of interest not only for purely descriptive reasons but also for the insights that it might provide into the particular variables that actually determine or mediate the variation of behavior with age. In this article, data are presented concerning the relationship between age and risk taking for an occupationally homogeneous sample of almost 1,500 mature adults. The form of the functional relationship between these two variables is examined, and possible mechanisms underlying it are discussed.

METHOD

Sample

The sample consists of 1,484 male managers, employed in over 200 corporations in a large variety of positions. All were engaged in management-development programs at the time of measurement, and the data were collected as part of a classroom demonstration. The age of the sample was approximately normally distributed with a mean of 39.34 yr., a standard deviation of 6.79 yr., and a range of 22–58 yr.

Instrument

Risk taking was measured by using a subset of the standard Kogan and Wallach (1964) choice-dilemma questionnaire. Items 1, 4, 7, 9, and 11 were presented in the given order. Briefly, these items consist of the following situations, all concerned with a safe, certain outcome versus a more desirable, riskier alternative:

1. A married engineer has to decide between a safe, secure job and a fledgling company that offers more responsibility and advancement.

---

[1] Requests for reprints should be sent to Victor H. Vroom, Graduate School of Industrial Administration, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pennsylvania 15213.

TABLE 1

ITEM INTERCORRELATION MATRIX

Item	1	2	3	4	5
1	1.000				
2	.056	1.000			
3	.087	.020	1.000		
4	−.001	.040	−.045	1.000	
5	.166	.083	−.001	.095	1.000

2. The captain of a college football team has to decide between a sure play to tie the arch rival, or a risky play to win.
3. A chess player is faced with a deceptive, risky maneuver that could expose him to certain defeat.
4. A prisoner of war is contemplating escape, knowing he will be executed if he is caught.
5. A research physicist has to decide between research topics that would be either plodding and sure or risky and uncertain.

Managers were tested in groups ranging in size from 20 to 60 participants. Each manager was given a booklet containing the five problems and an answer sheet on which to record his responses. He was asked to indicate the minimum probability of success that would warrant acceptance of the risky alternative for each problem. Responses to each item were to be chosen from the odd numbers 1, 3, 5, 7, and 9, representing the chances out of 10 that the risky alternative would be successful. If a respondent indicated that he would not choose the risky alternative regardless of its probabilities of success, his response was scored as 10.

In addition, each manager was asked to estimate whether the average participants in his group would respond with higher numbers to an item (i.e., they would be more conservative) or lower numbers (i.e., they would be riskier). If he believed that others would be more conservative than he was on an item, he indicated this with a plus following his own response; if he believed that others would be riskier, he indicated this with a minus following his own response. Those who believed that their response was at the average were instructed to record no sign.

Previous research (Hinds, 1962; Levinger & Schneider, 1969; Willems, 1969) has shown that *S*s tend to estimate themselves to be riskier than their peers. This finding has been interpreted as evidence for a cultural value placed on risk. People judge themselves as riskier than their peers because they like to think of themselves as riskier. The judgments of relative riskiness were obtained both to verify the previous findings of consistent overestimation and explore the relationship between age and a potential measure of the value placed on risk.

Following the completion of all five problems, the answer sheets were collected and the managers were divided into groups and asked to make group decisions on each problem. Only the individual responses are used in this article.

The use of the 5-item scale rather than the 12-item version in this research was dictated by time considerations. Undoubtedly, it had detrimental effects on the reliability of the scores obtained. There are two different bases for estimating reliability for the short version of the instrument. Kogan and Wallach (1964) and Wallach and Kogan (1961) present reliability estimates from several samples between the odd and even items of the 12-item measure. When these are weighed by their respective *N*s and stepped up by the Spearman-Brown formula, the result is an estimated reliability of the 12-item measure, $r_{11} = .62$. Decreasing the test length to five items leads to an expected reliability of the measure used in this study of $r_{22} = .4$. When the item intercorrelations shown in Table 1, based on data obtained from this sample are used as the basis for estimation, however, the reliability of the 5-item measure was estimated at only .2.

Either estimate indicates that the 5-item measure of risk taking is highly unreliable. This fact should greatly limit the size of any relationship between individual scores and age. The large number of cases, however, should still permit the determination of any relationships that do, in fact, exist.

RESULTS

Given the low intercorrelations among the items ($\bar{r} = .05$), independence of items was assumed and a median test constructed over all people-items to test the null hypothesis (i.e., the age groups were sampled from risk-equivalent populations). The *S*s ranged in age from 22 to 58 yr., and, therefore, constituted 37 age groups. The median response for all items was 5, and the number of item responses below 5 and the number of item responses ≥ 5 was counted for each age group. This test generates a value for χ^2 of 52.07 ($df = 36$, $.025 < p < .05$), so the null hypothesis is rejected and we conclude that there exists a significant age–riskiness relationship.

The median test treats age as a nominal variable. To ascertain the actual nature of the functional relationship between age and riskiness, we assume that the probability estimate scale is an interval scale and that the five items are to be weighted equally.

Given these assumptions, a riskiness index (R) was calculated for each S as the arithmetic mean of his responses. The higher the score, the less risky the person. The overall correlation between age (A) and riskiness (R) is:

$$r_{AR} = .0807 \quad (p < .001).$$

Both coefficients in the line regression line of

$$R = 3.895 + .0139A$$

are significant at beyond the .001 level.

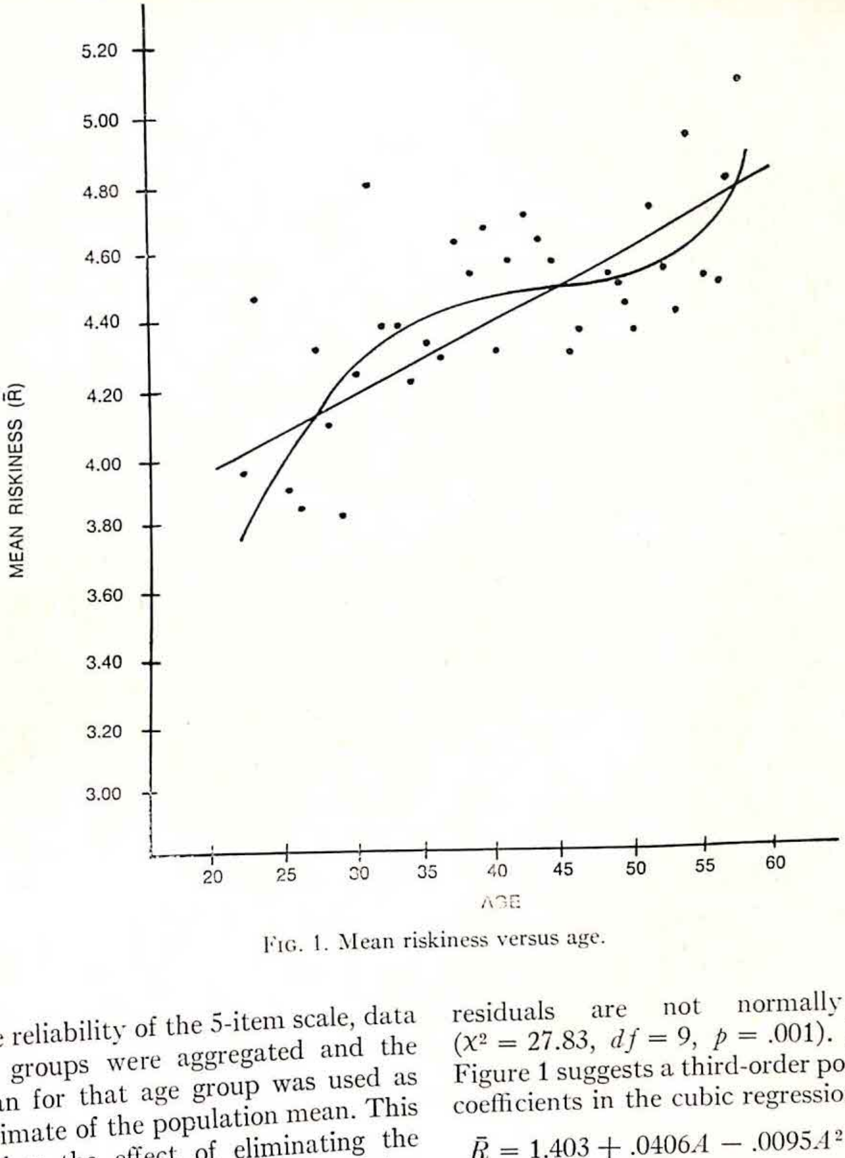


FIG. 1. Mean riskiness versus age.

Given the reliability of the 5-item scale, data within age groups were aggregated and the sample mean for that age group was used as the best estimate of the population mean. This essentially has the effect of eliminating the within-group variance, which is a combination of both error variance and true variance, and reduces the issue to a goodness-of-fit problem.

This correlation between age (A) and mean riskiness-age ($\bar{R}$) is:

$$r_{AR} = .64 \quad (t = 4.98, df = 35, p < .001).$$

Figure 1 shows the plot of this data. Also included in Figure 1 are two regression lines, a linear and a cubic.

The coefficients in the linear regression line of

$$\bar{R} = 3.575 + .0203A$$

are both significant beyond the .001 level. A chi-square test indicates, however, that the residuals are not normally distributed ($\chi^2 = 27.83,\ df = 9,\ p = .001$). Inspection of Figure 1 suggests a third-order polynomial. The coefficients in the cubic regression line of

$$\bar{R} = 1.403 + .0406A - .0095A^2 + .00007A^3$$

are all significant at the .05 level. A chi-square test indicates that the residuals are normally distributed ($\chi^2 = 8.2,\ df = 9,\ p = .5$).

These results indicate that risk taking as measured by the 5-item scale generally declines with age. From inspection of Figure 1 and the significance of coefficients in the cubic-regression line, it appears that the slope of the relationship is greatest in the age range 22–32 yr., flattens out in the age range 33–48 yr., and increases again in the age range 48–58 yr.

It remains to be seen, however, whether the relationship between age and risk taking applies to all five items or merely to a subset of

TABLE 2

CORRELATIONS BETWEEN AGE AND MEAN RESPONSE BY ITEM

Item	$r_{A\bar{R}}$	t	p
Engineer	.54	3.77	.001
Football	.04	.25	*ns*
Chess player	.47	3.15	.005
POW	.28	1.73	.05
Researcher	.44	2.93	.005

them. Table 2 shows the correlations between age and mean response to each of the five items. All correlations are positive and, with the exception of Item 5, are significant at the .05 level.

It is tempting to conclude from these findings that age and risk taking as measured by the Kogan and Wallach (1964) items are correlated in the total population of managers. It is possible, however, that the relationship may be artifactual due to particular and unique properties of our sample. The 1,484 managers were drawn from many different companies and worked in different functional areas within management. If, in our sample, managers in different functions or companies differed both in age and in risk taking, and if these differences did not exist in the total population, then the relationship between age and risk taking might be spurious.

TABLE 3

MEDIAN TEST, FUNCTIONAL AREA, AND RISKINESS

Functional area	No. below *Mdn.*	No. equal or above *Mdn.*	Total
Marketing	719	740	1459
Finance	351	372	723
Production	706	737	1443
Personnel	180	167	347
Research & development	273	304	577
General management	253	246	499
Engineering	418	441	859
Systems	58	82	140
Management staff	43	50	93
Other	514	560	1074
Totals	3515	3699	7214[a]

[a] Does not equal 7420 $\left(1484 \text{ People} \times 5 \frac{\text{Items}}{\text{Person}}\right)$ because of failure to respond or response between categories (i.e., even numbers). Since this is only 3% of the total response sample, this is hardly significant.

It is, of course, impossible to draw conclusions about the total population except insofar as it is revealed in our admittedly nonrandom sample. It is possible, however, to examine the existing data for relationships between uncontrolled variables and both age and risk taking.

Managers had categorized themselves as working within one of nine types of work: marketing, finance, production, personnel, research and development, engineering, systems, general management, and management staff. Nonreports and those that did not fit into any of these categories were placed in a tenth category termed "other."

Table 3 shows the number of item responses below the median and above or equal to the median of those 10 categories. There appears to be a slight tendency for those in general management and those in personnel to be slightly riskier than the total population, but a median test conducted to test the null hypothesis yields a value for χ^2 of 6.36 ($df = 9$, $p = .3$). The null hypothesis cannot be rejected, and there is no evidence for believing that the type of work performed is related to risk taking and the source of a spurious relationship between age and risk taking.

Differences in risk taking among the companies represented in the sample might also affect the relationship between age and risk taking. To test this possibility, three firms were selected that accounted for approximately two-thirds of the entire sample. They were: (*a*) Firm A, an integrated producer and fabricator of aluminum ($N = 251$); (*b*) Firm B, manufacturer of electrical and electronic equipment ($N = 410$); (*c*) Firm C, manufacturer of equipment for the generation, transmission, utilization, and control of electricity ($N = 341$).

A median test conducted to test the null hypothesis shows the value of χ^2 as 9.01 ($df = 2$, $p < .01$). There are significant differences in risk taking among these companies.

Table 4 shows the mean and variance on age and riskiness for each of the three companies. All of the differences in age are highly significant ($p < .001$) with the managers in Firm A being the oldest and Firm B being the youngest. The managers in Firm A have the highest scores on the risk-taking measure (i.e., are most conservative), and Firm C has the lowest score on that measure. The difference between Firm A and Firm B is significant ($t = 2.15$, $p < .025$) and in line with the age differences in the samples. The difference between Firms B and C approaches significance ($t = 1.47$, $p < .08$) but is in the reverse direction from what might be expected from the mean age of the samples.

Table 4 also shows the correlation between age and mean risk taking within each of three companies. The correlations are still positive, ranging from .61 to .42. The relationship between age and risk taking is not eliminated when company differences are eliminated.

It will be recalled that managers were also asked to indicate their perception of their relative riskiness by putting pluses and minuses beside their individual item response. This was done because of an assumed relationship between biases in *S*s' estimates and the amount of value that he attaches to risk.

Of 7,420 instances in which pluses and minuses could potentially have been recorded (1,484 Persons × 5 Items), such judgments were made 4,254 times or 57.3% of possible instances. These judgments are unevenly distributed between pluses and minuses with 2,712 pluses and 1,542 minuses. This is consistent with previous evidence that persons tend to think of themselves as riskier than their peers.

The critical question, however, concerns the relationship between these judgments and age. Once again using the lowest power test and the smallest number of assumptions about the data, a chi-square test was conducted to test the null hypothesis that the age groups constitute equivalent populations with regard to the distribution of pluses and minuses. This test generates a value for χ^2 of 80.11 ($df = 36$, $p < .001$), so the null hypothesis is rejected.

To explore the actual form of the functional relationship between age and the risk-value measure, an index was constructed for each person as follows:

$$Q = \frac{(\text{No. of pluses}) - (\text{No. of minuses})}{\text{Total no. of responses}}.$$

This index has a range of $+1$ to -1.

The correlation between age and the risk-value measure (r_{AQ}) is $-.13$ ($p < .001$). Since the measure Q, like R, may be expected to be highly unreliable, a correlation was computed between age and mean risk-value score ($r_{A\bar{Q}}$) and was found to be $-.76$ ($p < .001$). Figure 2 shows a plot of $\bar{Q}$ and age. Clearly, as age increases, the scores on the risk-value measure decrease.

This finding may be due to the fact that older persons place less value on thinking of themselves as risk takers. But, it is also possible that it is simply an artifact of their greater conservatism. An older person who states that he would take the risky alternative if it had 9 out of 10 chances of success has a more difficult time seeing himself as more risky than his peers than a younger person who states that he would take the risk if it had only 1 out of 10 chances of success. In effect, the measure Q may be highly correlated with R and the correlation r_{AQ} may be an artifact of the association of both A and Q with R.

The correlation r_{QR} was found to be $-.29$ ($p < .001$) and the correlation $r_{\bar{Q}\bar{R}}$ across age categories was found to be $-.70$ ($p < .001$). However, the partial correlation $r_{QA \cdot R}$ and $r_{\bar{Q}\bar{A} \cdot \bar{R}}$ were found to be $-.11$ ($p < .001$) and $-.56$ ($p < .001$), respectively. Clearly, age and the risk-value measure are correlated even when the influence of riskiness is partialled out.

The fact that age is correlated with estimates of relative riskiness in relation to peers even when scores on risk taking are partialed out has been attributed to the fact that younger persons place more value on risk taking than older persons. It is also possible that this finding reflects a widespread shared belief that younger managers are more risky than older ones. According to this explanation, a manager would be aware of any differences between him and other members of his group and would base his estimates of differences between his re-

TABLE 4

MEAN SCORES ON AGE AND RISK TAKING AND CORRELATIONS BETWEEN AGE AND RISK TAKING BY COMPANY

Firm	$\bar{A}$	S_A^2	$\bar{R}$	S_R^2	n	$r_{A\bar{R}}$	p
A	42.70	50.43	4.70	6.65	251	.61	.001
B	36.35	12.20	4.50	6.73	410	.42	.07
C	39.92	44.08	4.37	7.14	341	.49	.01

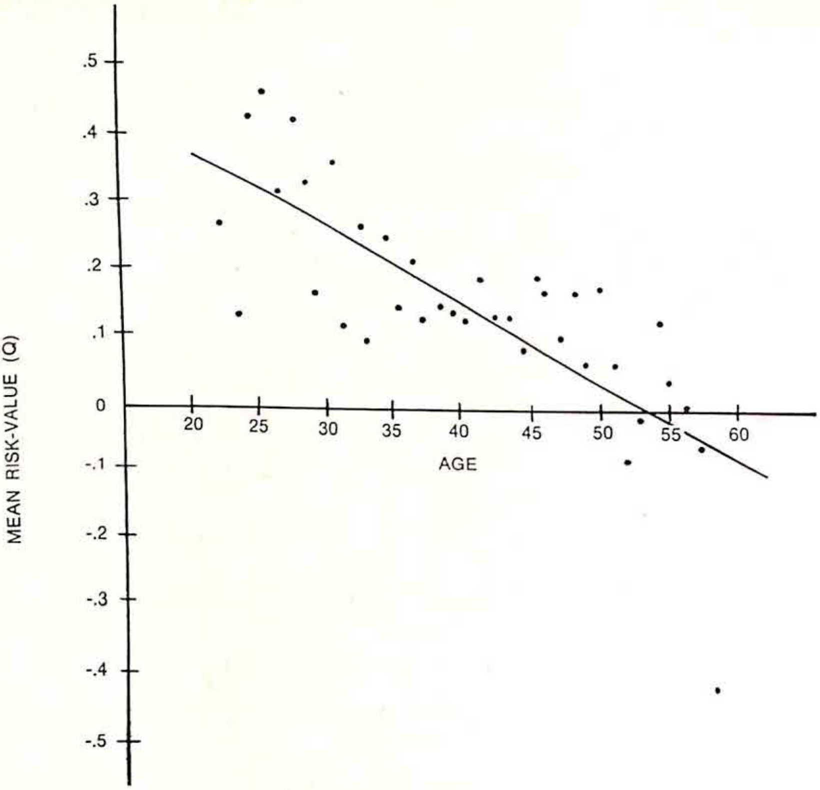


FIG. 2. Risk value versus age.

sponses and theirs on this knowledge. This explanation cannot be completely ruled out with these data. However, it does not explain why there should be an inordinately larger number of pluses to minuses. It seems slightly more parsimonious to assume that there is a widespread value placed on risk in American culture and that differences in the strength of this value are correlated with age.

DISCUSSION

The results show a significant relationship between age and measures of both risk taking and of the value placed on risk. The relationships are small in size (although highly significant) when individual scores are correlated with age but increase markedly when mean scores are used. Furthermore, the relationships are relatively stable across items and across companies.

What can be said about the process underlying this relationship? One possibility is that the regression line relating age to risk-taking constitutes a developmental track along which people move as they increase in age. As a person gets older, he gets married, has children, and acquires responsibilities to persons other than himself that militate against his taking risks that he might well have taken earlier in his development.

It is also possible that the explanation is not developmental at all, but cultural. The younger a person, the less likely he is to have experienced the depression, to have served in World War II, and the more likely he is to have been reared in an atmosphere of affluence and economic stability. According to this explanation, risk taking is a relatively stable attribute of personality that is learned fairly early in life with the sociocultural conditions prevalent at that time being instrumental in its determination. The points along the regression line do not constitute a track along which people move but rather, indications of sociocultural change. Each age group is moving along its own track, and that track is parallel to the horizontal axis.

These explanations are not, of course, mutually exclusive. Both processes of development and of sociocultural change may underly the relationship of age to risk taking. The critical question to be answered involves their relative contributions.

The data reported in the present article are not adequate for this purpose. A longitudinal investigation in which people are studied over time, particularly during the age ranges corresponding to the steepest slopes on the regression line would provide the most convincing answers.

But, the question can also be studied cross-sectionally. Let us assume that data of the general type reported here were collected from different samples of the same populations at intervals of 2 yr. over a 10-year period. If the developmental explanation accounted for the relationship between age and risk taking, there would be no change in the slope of the linear regression, points of inflection in nonlinear regression, intercepts, or mean scores on risk taking. If the sociocultural explanation were correct, the slope might remain the same, but all other parameters would change systematically over time.

REFERENCES

HINDS, W. C. Individual and group decisions in gambling situations. Unpublished master's thesis. School of Industrial Management, Massachusetts Institute of Technology, 1962.

KOGAN, N., & WALLACH, M. A. *Risk-taking: A study in cognition and personality*. New York: Holt, Rinehart & Winston, 1964.

KOGAN, N., & WALLACH, M. A. Risk-taking as a function of the situation, the person, and the group. In *New directions in Psychology III*. New York: Holt, Rinehart & Winston, 1967.

LEVINGER, G., & SCHNEIDER, D. J. Test of the "risk is a value" hypothesis. *Journal of Personal Social Psychology*, 1969, **11**, 165–169.

WALLACH, M. A., & KOGAN, N. Aspects of judgment and decision making: Interrelationships and changes with age. *Behavioral Science*, 1961, **6**, 23–36.

WILLEMS, E. P. Risk is a value. *Psychological Reports*, 1969, **24**, 81–82.

WOHLWILL, J. F. The age variable in psychological research. *Psychological Review*, 1970, **77**, 49–64.

(Received June 23, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 406–413

MOTIVATIONAL TYPE AND THE SATISFACTION-PERFORMANCE RELATIONSHIP [1]

FRANK J. LANDY [2]

Pennsylvania State University

Recent work in the area of job satisfaction and work motivation has emphasized the importance of individual differences. The logic of subgrouping was combined with a consideration of individual differences in the motivation to work in an attempt to shed some light on the satisfaction-performance relationship. Professional engineers were subgrouped on three motivational dimensions. Motivational type was tested as a moderator of the relationship between five factor analytically derived satisfaction dimensions and six performance dimensions. Two of the emerging types showed moderator characteristics.

There has been considerable discussion of the merits of subgrouping and typal analysis over the past few years (Owens, 1968). The purpose of the present research was to determine if such an approach might lead to a more efficient description of the relationship between job satisfaction and job performance than has been supplied to this point in the literature.

The study of various reinforcers as well as the individuality of satisfaction patterns set the stage for the introduction of subgroup or typal logic. Two cornerstones of this emerging concern for individual differences were Shaffer's (1953) monograph on need satisfaction and Ahlberg and Honey's (1950) study of government scientists. Although both of these works pointed to the importance of individual differences in the consideration of job satisfaction, they did not concern themselves with the identification and description of subgroups. More recent and considerably more direct attempts to introduce the typal approach in the analysis of the motivation and satisfaction constructs have come from Katzell, Barrett, and Parker (1961); Dunnette, Campbell, and Hakel (1967); and Graen, Dawis, and Weiss (1968).

The Graen, Dawis, and Weiss (1968) study introduced a unique strategy insofar as they not only identified motivational types, but they also used these types as moderators of the relationship between need-reinforcer correspondence and job satisfaction. They found that need type did, in fact, act as a moderator. It is important to note that type acted as a multivariate moderator, that is, the pattern of needs of the individual was of maximal importance.

The present research is based on the notion that motivational type or subgroup might act as a multivariate moderator of the relationship between job satisfaction and performance. The logic of the hypothesized motivational moderator is closely related to the consideration of satisfaction and performance as collinear variables tied together by rewards. The problem becomes one of defining "rewards." The importance of the concept of individual differences, as previously discussed, is that the nature of a reward varies from person to person. The identification of motivational types or subgroups was intended to control, in a gross sense, the salience of reinforcers in the individual's environment. People having similar reward systems (i.e., the same reward salient for each) should, given similar rewarding systems, be more satisfied and perform at higher levels (if rewards are, in fact, tied to performance) when their own particular needs are satisfied.

The present study is the first step in describing the various relationships between specific need patterns, satisfaction with specific aspects of the work, and specific job behaviors, as outlined above. Its primary intent was to demonstrate that the pattern of individual needs modifies the relationship between job satisfaction and performance.

---

[1] This research was partially supported by a grant from the United States Department of Labor and the central fund for research at Pennsylvania State University.

[2] Requests for reprints should be sent to Frank J. Landy, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802.

Method

The *S*s chosen for the study were 175 engineers with no supervisory responsibilities who came from six different organizations. Five of these organizations were commercial firms with engineering subsections, while the sixth was an engineering consulting firm. All data were gathered in group meeting rather than by mail.

There were four broad classes of variables in the study: (*a*) satisfaction variables, (*b*) motivation variables, (*c*) performance variables, and (*d*) work history and biographical information. These classes are described separately.

Satisfaction variables. A questionnaire constructed by Elbert (1966), which will be referred to as the satisfaction inventory (SI), was chosen as a measure of the satisfaction components in one's job. The questionnaire was developed by writing statements for each of 21 need categories identified in the literature. Internal consistency analysis was used to identify the five best statements in each category. The inventory was administered to male college graduates employed for 3 yr. or less. A 7-point response scale of satisfaction was used signifying the degree to which the item was perceived to be present in the job. A principal axis factor analysis with a Varimax rotation identified 12 need satisfaction factors accounting for 98% of the common variance.

For the present study, a pilot was done to determine the adequacy of the SI for a relatively homogeneous occupational group (professional engineers). Consequently, the SI was mailed to engineers employed by three different organizations. Of the 250 questionnaires that were sent out, 176 were returned. The pilot study indicated that the SI was suitable for use in the collection of satisfaction data. The same major factors appeared (in terms of percentage of variance accounted for), and they were defined by the same items as in Elbert's (1966) analysis. The factors and items are more fully described in the Results section.

Motivation variables. These consisted of peer ratings of the motivation to work on seven dimensions. The scales are labeled as follows: Professional Identification, Job Curiosity, Team Attitude, Task Concentration, Independence/Self-Starter, Persistence, and Organizational Identification. A description of these scales and their development has recently been published (Landy & Guion, 1970).

Performance variables. Sprecher (1965) has developed a series of 33 performance scales for use by supervisors in rating professional engineers. The scales are considered representative of five broad factors: divergent production, convergent production, responding and doing, evaluation, and cognition. From the 33 scales, 6 were selected for use in the present study on the basis of statistical independence and general suitability. The scales were: Problem Weighting, Communication of Results, Use of Personnel Resources, Personal Ethics, Problem Analysis, and Valuable Results. Performance and motivation of *S* were rated by co-workers selected by *S*.

Additional data. The *S*s were also asked to supply biographical information and work history information, as well as to complete the Bass Orientation Inventory (ORI; Bass, 1962). The Bass ORI yields three scores: Self-Orientation, People- or Interaction-Orientation, and Task Orientation.

Results

Satisfaction

The SI was cluster analyzed using Tryon's key cluster analysis as found in the BC-TRY computer system of cluster and factor analysis (Tryon & Bailey, 1966). Cluster analysis of the satisfaction inventory yielded five clusters accounting for 97% of the variance of the raw correlation matrix and 96% of the communality. The five clusters were: Advancement, Ethical Principles, Creativity, Pay, and Working Conditions. The items defining the clusters and their loadings can be found in Table 1. The intercorrelations of the clusters, the cluster reliabilities, and the percentage of variance accounted for by each cluster can be found in Table 2.

Motivation

Rather than assume the independence of motivational dimensions for typing (and possibly introduce an error of overrepresentation of a particular motivational domain), a factor analysis of the motivation ratings was done. A principal factors solution was applied with both a Quartimax and Varimax rotation. This analysis indicated that the three most independent dimensions for typing (i.e., those which had the greatest potential for yielding discriminably different types) were Professional Identification, Team Attitude, and Task Concentration.

For the motivational typing or subgrouping, a minimum within-cluster variation technique was used. The procedure is one of the components of the BC-TRY computer system and is known as O-TYPE. The process is one of clustering objects in multidimensional space. The dimensionality in this case was three (the three motivational dimensions). As a function of standard scores, each *S* was classified as high, moderate, or low on each of the motivation dimensions (high = +1 *SD* and higher,

TABLE 1

CLUSTER DEFINERS AND LOADINGS

Cluster	Loading	Item
Advancement		
	.91	The job provides an opportunity for advancement.
	.89	You have an opportunity to get ahead.
	.84	There is a chance to move one job higher.
	.66	You advance.
Ethical Principles		
	.91	The things you do are morally right.
	.91	Your ethical standards are maintained.
	.79	Your work agrees with your ethical principles.
Creativity		
	.82	Your creative ability is useful on the job.
	.86	You can be imaginative and think up unique ways of doing things.
	.79	You can find new ways to do things.
	.57	You have the freedom to use your own judgment.
Pay		
	.82	You are paid enough for the work you do.
	.81	You are able to get the pay increases you feel you deserve.
	.77	You are satisfied with the length of time between pay increases.
	.76	You feel secure with the amount of money you earn.
Working Conditions		
	.95	There are good physical working conditions.
	.71	You have pleasant work surroundings.
	.61	Working conditions, that is, heating, lighting, etc. are adequate.

Note.—Oblique factor coefficients.

moderate = +1 to −1 *SD*, low = −1 *SD* and lower). This allows for 27 (or 3^3) a priori types. This score sector space is a starting point for an interative clustering procedure based on Euclidian distances.

One of the problems involved in the choice of a typing strategy concerns the model chosen. The deterministic model is stated in terms of the ideal case in which no provision is made for chance or error variance in the model itself, and every object must be allocated to a specific subgroup. Since it was unreasonable to assume that the motivational domain was completely represented (as manifested in the types) by three dimensions, a probabilistic model was chosen for typing. This strategy allows for variation in descriptor variables and the persons described, and not all persons need belong to one of the subgroups found (Lorr, 1968).

TABLE 2

CLUSTER INTERCORRELATIONS AND RELIABILITIES

Cluster	1	2	3	4	5
1. Advancement	90	23	38	34	22
2. Ethical Principles		91	27	14	28
3. Creativity			87	−05	14
4. Pay				89	20
5. Working Conditions					81
% h^2	32	24	20	13	07

Note.—Reliabilities are Spearman-Brown reliabilities (see Tryon [1957] and Ghiselli [1964]). Decimals omitted.

The subgroupings resulted in 13 distinctive types comprising 171 *S*s. The remaining four *S*s were considered unique or isolates, a phenomenon consonant with the probabilistic model used. The motivation types, their standard score patterns, and sizes are presented in Table 3.

Another component of the BC-TRY computer system known as 4CAST enables one to test the relationship between types and several other variables of interest (in this case work history variables and Bass ORI scores). The component is appealing because it employs a Monte Carlo sampling routine that allows one to implement the central limit theorem and thereby treat each test as independent. Briefly, the procedure draws a large number of random samples (e.g., 3,000) from your overall sample. Each of these samples is of the same size as the type you are examining. The mean of the variable of interest for the type is then superimposed on the distribution of means for the random samples. Statements can then be made about the significance of the difference between the type mean and the overall sample mean. On the basis of this kind of analysis, the following statements can be made about the motivation types: Type 2—significantly higher on Self-Orientation ($p < .05$), significantly lower on Task-Orientation ($p < .05$); Type 3—significantly higher on People-Orientation ($p < .05$); Type 4—significantly lower on People-Orientation ($p < .05$), significantly higher on Task-Orientation ($p < .001$); Type 5—significantly fewer previous jobs ($p < .01$); Type 6—significantly higher on People-Orientation ($p < .001$); Type 7—significantly lower on People-Orientation ($p < .01$), significantly higher on Task-Orientation ($p < .05$); Type 9—significantly greater number of years past degree ($p < .05$).

While the independence of the types is assured by the clustering process, unless one has a large sample to start with, many of the emerging types or subgroups are too small for stable inferences of the form required by correlational forms of moderator analysis. While it is possible to combine the types in a specified hierarchical order to achieve larger cell sizes, it was felt that more information would be lost in the combining process than would be gained by larger *N*s. Consequently, a compromise was made. It was arbitrarily decided to continue the moderator analysis with types comprising 20 or more *S*s. This was felt to be a more adequate strategy than combining because one can take sample size into account and adjust accordingly in significance tests. Similar problems were encountered in a motivational typing study done by Graen, Dawis, and Weiss (1968). This will probably continue to be a problem as long as models are used that assume a multivariate-normal surface.

Two of the four types showed moderator characteristics. An examination of Tables 4 and 5 shows that the full-sample correlations are generally of zero order, while the moderated correlations describe significant relationships between satisfaction and performance. The number of significant correlations in the moderated matrix is sufficient to indicate that their appearance is not solely due to chance.

TABLE 3

MOTIVATION TYPE SCORE PATTERNS

Type	Motivation dimension			*N*
	Professional Identification	Team Attitude	Task Concentration	
1	32.9	29.0	44.8	6
2	38.9	46.2	26.6	7
3	28.7	50.5	50.7	8
4	50.9	31.6	54.0	10
5	47.2	53.4	38.5	21
6	45.6	46.6	49.2	29
7	49.6	52.3	57.9	21
8	51.6	57.5	50.6	21
9	46.0	61.2	63.0	10
10	68.9	54.1	38.4	5
11	62.1	45.4	51.4	11
12	64.5	55.9	63.5	10
13	62.6	58.0	53.2	12

Note.—All scores appear in standard score form ($\bar{X} = 50$, $SD = 10$).

There is an interesting statistical point that arises from the moderator analysis. When you are dealing with a matrix of correlation coefficients rather than a single value, what should be considered as evidence for the presence of moderator effects? The first and most obvious is simply to look at the increase in the number of significant correlations and compare that to levels you might expect by chance. Since the overall matrices presented in Tables 4 and 5 are essentially of zero order, the differences between the moderated matrices and the overall matrix are fairly obvious. On the basis of this kind of approach, there does seem to be some justification for accepting the moderator characteristics of motivational type.

But since we are directing our attention to the grosser question of the relationship between satisfaction and performance (the ex-

TABLE 4

INTERCORRELATIONS OF SATISFACTION CLUSTER SCORES WITH PERFORMANCE VARIABLES, AND MODERATED CORRELATIONS

Variable	Satisfaction dimension				
	Advancement	Ethical Principles	Creativity	Pay	Working Conditions
Problem Weighting					
Intercorrelation	04	05	−02	09	06
Moderated correlation	53*	50*	07	06	31
Communicates Results					
Intercorrelation	02	01	04	15	01
Moderated correlation	27	13	13	54**	18
Personnel Resources					
Intercorrelation	03	07	06	03	01
Moderated correlation	60**	30	33	−12	18
Personal Ethics					
Intercorrelation	−11	−15	−05	02	01
Moderated correlation	30	09	09	00	05
Valuable Results					
Intercorrelation	−04	03	05	02	−15
Moderated correlation	35	42	27	−27	40
Problem Analyzing					
Intercorrelation	−06	12	14	04	02
Moderated correlation	23	45*	22	−02	50*

Note.—$N = 175$; Type 5 moderated correlation, $N = 21$; decimals have been omitted.
* $p < .05$ (two-tailed test).
** $p < .01$ (two-tailed test).

istence but not necessarily the degree and kind), perhaps a different kind of analysis is called for—one of a grosser type. One strategy might be to do an r to z' transformation on all values in the moderated and the overall matrix, respectively, and compute the significance of the difference between the average value for each of the matrices.

A related strategy might be to compare the root mean squares of the overall and moderated matrices, respectively. A modification of the nonparametric sign test might also be appropriate.

A final, and slightly more elaborate solution might be one involving canonical correlation. In the present case, the canonical correlation would take the form of a statement of the relationship between a linear composite of the satisfaction cluster scores and the linear composite of the performance scores. Due to the peculiar problems of interpreting overlapping variance, only the first canonical vector would be interpretable. Comparisons could be made between variance accounted for by the first vector of the overall matrix as compared to the variance accounted for by the first vector of the moderated matrix.

The point is not that there are lots of ways to look at the data; the point is that one of these strategies is probably more appropriate than the others for the description of moderating effects in a matrix of correlations.

DISCUSSION

It does seem as if motivational type acted as a moderator of the relationship between satisfaction and performance. In the present context, motivational type can be considered a pattern of energy expenditure. At first glance this might appear far removed from the more orthodox notions of motivation advanced in recent years, that is, motivation as a need system. The present conception of a motivational system was arrived at partly on methodological and partly on theoretical grounds.

Most researchers in this area have been struck, at one time or another, with the possibility that they may be simply capitalizing

TABLE 5

INTERCORRELATIONS OF SATISFACTION CLUSTER SCORES WITH PERFORMANCE VARIABLES, AND MODERATED CORRELATIONS

Variable	Satisfaction dimension				
	Advancement	Ethical Principles	Creativity	Pay	Working Conditions
Problem Weighting					
Intercorrelation	04	05	−02	09	06
Moderated correlation	26	−08	−16	−15	−13
Communicates Results					
Intercorrelation	02	01	04	15	01
Moderated correlation	29	−18	−52*	−21	−30
Personnel Resources					
Intercorrelation	03	07	06	03	01
Moderated correlation	01	07	−20	−13	−08
Personal Ethics					
Intercorrelation	−11	−15	−05	02	01
Moderated correlation	50*	−39	−55**	−04	−33
Valuable Results					
Intercorrelation	−04	03	05	02	−15
Moderated correlation	35	−49*	−22	−13	−53*
Problem Analyzing					
Intercorrelation	−06	12	14	04	02
Moderated correlation	50*	−32	−32	−20	−26

Note.—$N = 175$; Type 8 moderated correlation, $N = 21$; decimals have been omitted.
* $p = .05$ (two-tailed test).
** $p = .01$ (two-tailed test).

on response set when they correlate self-reported needs with self-reported satisfaction. I have also been struck with this possibility. The present research strategy has been to infer motivation from energy expenditure. This was made possible by the nature of the peer ratings. Co-workers were instructed to rate *Ss* on the degree of energy that they expended in each of the specified motivational dimensions. The development of the motivational scales makes it reasonable to assume that there is a fairly large behavioral component incorporated since interrater agreements reached generally acceptable levels. The logic is very simply that, given the opportunity, people will engage in activities that they find personally rewarding. Most studies have asked *Ss*, in one form or another, what conditions were necessary to produce movement. The present study considers observed movement as an index of motives.

The common ground between the more traditional view of motivation and the present approach is reward system and reinforcement history. These concepts will ultimately tie together need systems as predispositions for behavior and energy expenditure viewed as an overt behavior system.

Interestingly enough, the dimensions chosen for motivational typing turn out to be similar to dimensions suggested recently by Alderfer (1969) as basic dimensions of human needs. Alderfer proposes a threefold conceptualization of human needs comprising existence, relatedness, and growth. Existence needs seem to be characterized by a concern for satisfaction of personal needs not directly related to other people or the organization. Relatedness needs involve relationships with significant others. The third category, growth needs, ". . . include all the needs which involve a person making creative or productive effects on himself and the environment [p. 146]."

In the present study, Professional Identification seems closest to the proposed existence need, though Alderfer (1969) might contend that it is more appropriate to growth needs. The relationship between relatedness needs and the dimension Team Attitude is fairly obvious. The Task Concentration dimension effectively couples the individual with his work environment, as suggested in Alderfer's elaboration of growth needs.

While a description of motivational types present in a sample of *S*s might be of interest in its own right, the real value of such a procedure lies in relating these types to variables other than those used in the typing process. The primary example of such an approach in the present data is the moderator analysis, relating the types to the performance-satisfaction relationship. Another attempt at anchoring the types was through the use of the Bass (1962) ORI scores and the work history data. The relationships of the ORI scores to the types are not very clear and are mentioned in the hope of generating hypotheses rather than answering questions. There seems to be a loose relationship between ORI Task Orientation and the motivational dimension of Task Concentration. Those subgroups whose Task Orientation scores were significantly higher than the rest of the sample also had Task Concentration scores higher than the mean of the sample while the one group that had a Task Orientation score significantly lower than the rest of the sample had a Task Concentration score lower than the mean of the sample. This would seem to indicate some gross positive relationship between self-reported motivational direction and peer ratings of energy expended in the same domain. The same relationship does not seem to hold for ORI Interaction scores and Team Attitude. This is probably due to the fact that Interaction Orientation involves seeking out companionship while Team Attitude deals primarily with helping and working cooperatively with other people. This may indicate that seeking companionship is not a sufficient condition for being a good co-worker. Finally, it is interesting to note that the only group having a Self-Orientation score significantly higher than the rest of the sample was also rated below the mean on all three motivation dimensions.

The present data suggest that there is some value to breaking out of the framework of measuring motivation through self-report devices. It appears that knowing the energy expenditure patterns of professional engineers provides valuable information about satisfaction-performance relationships. As is obvious from the data analysis, the satisfaction-performance problem is not solved. Only two of the motivational types yielded relevant information (i.e., showed moderator characteristics), but this is completely acceptable under the notion of differential predictability. It is entirely possible that if the motivation typing were expanded to take into account more than three dimensions, predictive efficiency might have been increased.

There are two important things to keep in mind concerning the motivational types. Of primary importance is the fact that these subgroups represent multivariate combinations, taking into account the complexity of motivational systems. There seems to be a temptation in certain kinds of research to identify a subgroup, look at the scores of a series of single variables for the people in this subgroup, assume that all the people in the subgroup have identical amounts of the traits represented by these single variables, assume that these variables combine in exactly the same way for all people in the subgroup, and talk about "types" of people. A prime example of this logic appears in a recent book by Hackman (1969, p. 54) on work motivation. This strategy is occasionally dangerous and almost always misleading.

Of further importance in considering the present motivational types is the fact that they represent clusters of real people, that is, we can expect to find motivational dimensions combining in this manner. Again, there is a tendency to take normative data, partition the continua arbitrarily, combine these partitions, and talk about the resulting "types" as if they were actually represented in the data. This logic is, again, represented in Hackman's book (1969, p. 149). A related error is the interpretation of a vector of factor loadings derived from a factor analysis of variables as if it described subgroups of people present in the data. An example of this strategy appears in an article by Katzell, Barrett, and Parker (1961). The error in the manipulations previously described is that they should constitute the beginning of a typological analysis, not the end! The fact that disjunctive or incompatible patterns of scores can exist, at least in the realm of ability dimensions, has been documented by Tryon and Bailey (1970).

REFERENCES

AHLBERG, C. D., & HONEY, J. C. Attitudes of scientists and engineers about their government employment. Unpublished manuscript, Syracuse University, Maxwell Graduate School, 1950.

ALDERFER, C. An empirical test of a new theory of human needs. *Organizational Behavior and Human Performance,* 1969, **4**, 142–175.

BASS, B. M. *The orientation inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1962.

DUNNETTE, M., CAMPBELL, J. P., & HAKEL, M. Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Organizational Behavior and Human Performance,* 1967, **2**, 143–174.

ELBERT, A. J. Factor analysis of work relevant need statements in two populations. Unpublished master's thesis. Bowling Green State University, 1966.

GHISELLI, E. E. *Theory of psychological measurement.* New York: McGraw-Hill, 1964.

GRAEN, G. B., DAWIS, R., & WEISS, D. J. Need type and job satisfaction among industrial scientists. *Journal of Applied Psychology,* 1968, **52**, 286–289.

HACKMAN, R. *The motivated working adult.* New York: American Management Association, 1969.

KATZELL, R., BARRETT, R. S., & PARKER, T. S. Job satisfaction, job performance, and situational characteristics. *Journal of Applied Psychology,* 1961, **45**, 65–72.

LANDY, F. J., & GUION, R. M. The development of scales for the measurement of work motivation. *Organizational Behavior and Human Performance,* 1970, **5**, 193–203.

LORR, M. A review and classification of typological procedures. Paper presented at the meeting of the American Psychological Association, San Francisco, September 1968.

OWENS, W. Toward one discipline of scientific psychology. *American Psychologist,* 1968, **23**, 782–785.

SCHAFFER, R. H. Job satisfaction as related to need satisfaction in work. *Psychological Monographs,* 1953, **67** (14, Whole No. 364).

SPRECHER, T. A. Dimensions of engineers job performance. Paper presented at the meeting of the American Psychological Association, Chicago, September 1965.

TRYON, R. C. Reliability and behavior domain validity-reformulation and historical critique. *Psychological Bulletin,* 1957, **54**, 229–249.

TRYON, R. C., & BAILEY, D. E. The BC-TRY computer system of cluster and factor analysis. *Multivariate Behavioral Research,* 1966, **1**, 91–111.

TRYON, R. C., & BAILEY, D. E. *Cluster analysis.* New York: McGraw-Hill, 1970.

(Received June 1, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 414–421

TASK ENJOYMENT AND TASK PERSEVERANCE IN RELATION TO TASK SUCCESS AND SELF-ESTEEM

SKIPTON LEONARD AND JOSEPH WEITZ[1]

New York University

An attempt was made to use self-esteem as a moderator for the relationship between success-failure and task enjoyment. It was hypothesized that success-failure would be related to enjoyment for a high-self-esteem group but not for a low-self-esteem group. This was not the case. Success and failure were related to task enjoyment regardless of the self-esteem of S as determined by typical measures of this variable. It was further found that a possible behavioral measure of task enjoyment (perseverance on a task) was not highly related to a rating of task enjoyment. This gives one pause concerning the meaning of ratings of enjoyment.

The relationship between cognitive balance and behavior has been examined in a variety of contexts in recent years. A self-consistency mechanism has been proposed by Korman (1968) in his discussions concerning the relationships between task achievement, self-esteem, and task satisfaction. This self-consistency model is a natural extension of cognitive integration or balance theories as proposed by Lecky (1945), Festinger (1957), and Vroom (1964) and reflects a sophistication that has developed from recent job satisfaction studies (Locke, 1965; 1966; c.f. Argyris, 1964; Dunnette, Campbell, & Hakel, 1967) and studies in social influence theory (Deutsch & Gerard, 1955).

The essential assumption of the self-consistency model was stated as follows: "All other things being equal, individuals will a) choose, b) adjust their performance to, and c) find satisfying those behavioral roles which maximize their sense of cognitive balance or consistency [Korman, 1968, p. 485]." It is within this framework that we shall investigate task satisfaction as it relates to success and failure and to self-esteem.

One's self-esteem or self-assurance can be seen as having two, conceptually semiindependent components. General or chronic self-esteem is related to an individual's self-evaluation of his overall capabilities, and, as such, might be thought of as a personality characteristic that is relatively stable over time. Another component of self-esteem may be considered as task specific. This corresponds to an individual's perceived ability in certain specific areas, and its relation to general or chronic self-esteem is somewhat unpredictable. For instance, an individual may think highly of his ability to perform well in his chosen work and to handle most problems that might come up in this area and yet have a low self-appraisal of his ability in certain other specific areas.

If one's cognitive makeup with respect to self-esteem can be viewed in these terms, then it becomes apparent that cognitive balance need not be an all-or-none operation. Certain behaviors may be inconsistent for one part of one's cognitive structure and consistent for another part.

Using a self-consistency or balance model, it was hypothesized that success or failure on a task would be related to enjoyment of that task for individuals having high-self-esteem but not for those having low-self-esteem.

Present Experimental Paradigm

In the present experiment, *S*s were given a time limit to solve a puzzle involving spatial relations. Those who solved the puzzle in the time period alotted were designated successes while those who were not able to solve the puzzle were designated failures. Success or failure in this situation was self-evident to *S*s, and no manipulations were required to form success or failure conditions. (This fits the experimental paradigm suggested by

---

[1] Requests for reprints should be sent to Joseph Weitz, Department of Psychology, New York University, 21 Washington Place, Room 311, New York, New York 10003.

Weitz [1967].) The *S*s were then asked to evaluate their enjoyment of the task and also how appropriate they thought the task was for measuring intelligence and for studying problem-solving behavior.

At this point in the experiment, *S*s were given another spatial relations puzzle to solve but were given as much time to solve the puzzle as they wished. They were also informed of a third task that might be started at any time if they could not solve the second puzzle. The number of minutes that *S*s worked on the second puzzle before giving up and trying the third task was a major dependent variable of interest in this experiment. After each *S* had solved or had given up on the second task, he was asked again to evaluate his enjoyment of the task, how appropriate he felt the task was for measuring intelligence, and how appropriate he felt the task was for studying problem-solving behavior. A third puzzle was given to *S*s to work on, but no measures were taken since it was only included to keep the experiment plausible. This design is similar to, but not a replication of, Korman's (1968, Experiment 1) study. A different basic task was used, there was no manipulation by *E* of success and failure, and certain behavioral measures were included.

METHOD

Subjects

The *S*s for the present experiment were recruited as usual from introductory psychology classes. Students in these classes are required to serve as *S*s in psychology department-approved experiments. Sixty-eight males and 80 females were obtained in this manner.

Apparatus and Measurement Instruments

Three puzzles were designed for the present experiment. These puzzles were constructed by cutting 12-in. square masonite panels into four to six pieces that could be assembled to reconstruct the square. The approximate difficulty of each puzzle was assessed during pretesting sessions where it was determined that Puzzle 1 was, in general, an easier puzzle to solve than Puzzles 2 or 3.

Three measures of self-esteem were employed in the present design. The Ghiselli Self-Descriptive Inventory (undated) consists of 64 pairs of personally descriptive adjectives and yields a self-assurance scale that has been used in a number of studies involving self-esteem (Gavin, 1969; Korman, 1968; Schein, 1969). Half of the pairs are adjectives of approximately equal positive social desirability, and the other half are pairs of adjectives with approximately equal negative social desirability.

The manual for the Ghiselli Self-Descriptive Inventory (undated) attempts to define the concept of self-assurance as follows:

> Self-assurance refers to the extent to which the individual perceives himself to be effective in dealing with the problems that confront him. There are those persons who see themselves as being sound in judgment and able to cope with almost any situation, whereas others think of themselves as slow to grasp things, making many mistakes, and being generally inept (Ghiselli, undated, p. 9).

Korman (1968) reports data that suggest that this empirically based scale possesses moderate construct validity.

The Trait Evaluation Index (Nelson, 1968) yields self-confidence and personal adequacy scales. The Trait Evaluation Index consists of 125 triads of personally descriptive adjectives that have been matched for positive social desirability. The respondent was to judge which adjective of the triad was most descriptive of him and which least descriptive. Two examples of the triads are shown below:

A. Sportsmanlike
B. Quick
C. Self-reliant

A. Independent
B. Earnest
C. Self-assured

The manual for the Trait Evaluation Index does not attempt to define the concepts of self-confidence or personal adequacy and these concepts, therefore, can only be defined operationally in terms of those words keyed to those scales. There are data reported in the manual that indicate some construct validity for the Trait Evaluation Index.

A 5-point rating scale was constructed to assess *S*s' evaluation of the first two tasks. The *S*s were asked to respond to the following three questions after each of the first two tasks: (*a*) How enjoyable did you find this puzzle to be? (*b*) How appropriate do you feel this problem is for measuring intelligence? (*c*) How appropriate do you feel this problem is for studying problem-solving behavior?

Procedure

The *S*s were instructed to complete a short biographical inventory and two personality inventories. During this time, *S* was not supervised by *E*. Upon completion of these inventories, *E* read a short description of the purpose of the experiment and instructions for the first task. The general description and instructions given *S* for the first task were as follows:

> As a subject in this experiment you are helping to develop a new intelligence or aptitude test which, hopefully, has less cultural bias than the older intelligence tests.
>
> You will be asked to work on a number of problems which we believe to be related to intelligence. The results may have important implications

TABLE 1

INTERCORRELATION MATRIX FOR THE INDEPENDENT VARIABLES

Variable	S-F (Task 1) (1 = failure, 2 = success)	S-F (Task 2) (1 = failure, 2 = success)	Sex (1 = female, 2 = male)	SA (1–48) (low) (high)	SC (1–45) (low) (high)	PA (1–39) (low) (high)
S-F (Task 1)						
S-F (Task 2)	.22					
Sex	.19	.17				
SA	.06	.07	.24			
SC	.06	−.02	.17	.30		
PA	.07	.06	.21	.26	.25	

Note.—The direction in which variables were scored is given below each variable. Pearson *r*s > .160 are significant at the $.05_2$ level and *r*s > .209 are significant at the $.01_2$ level for $df = 146$. Abbreviations: S-F = success-failure, SA = self-assurance, SC = self-confidence, PA = personal adequacy. $N = 148$ (68 males, 80 females).

> in the intelligence-testing field. Your full cooperation, therefore, will be most appreciated.
>
> The masonite pieces which I have placed before you on the desk form a perfect twelve-inch square when properly assembled. One side of each piece has a very rough surface. This rough surface should always be kept face down. There is only one solution to this puzzle and that solution requires that no piece has its rough surface showing. Are there any questions? You will be given two minutes to solve this puzzle. Are you ready? You may begin.

Each *S* was then allowed 2 min. to assemble a four-piece puzzle. After the 2 min. had elapsed, *S* was instructed to rate his enjoyment of the task, appropriateness of the task for measuring intelligence, and appropriateness of the task for studying problem-solving behavior.

Upon completion of these ratings, *E* read the instructions for the second task.

> The instructions for the second problem are the same as for the first problem. In other words, it forms a perfect twelve-inch square. The puzzle is different from the first, and you should have more difficulty in solving it. You will be given as much time to solve this puzzle as you wish, but since it is a difficult puzzle, you may start on the third problem at any time after you have started the second puzzle. This third problem may be a little easier than the second puzzle. If you wish to start the third problem, simply tell me so and I will start you working on it.

After the *S* had solved the second task, or more frequently, had decided to start the third task, he was instructed to rate his enjoyment of the second task, the appropriateness of the second task for measuring intelligence, and the appropriateness of the second task for studying problem-solving behavior.

It should be noted that the words "quit," "fail," or "success" were never used by *E*. No words of praise, disappointment, or general comment were expressed after any of the puzzles. If *S* wished to leave the second puzzle and try to solve the third puzzle, he did so at his own initiative.

Experimental Predictions

The predictions for ratings of enjoyment of the first task were that within the high-self-esteem group those who succeeded would enjoy the task more than those who failed. No such prediction was made for the low-self-esteem group. These predictions are similar to those of Korman (1968).

The rationale for the prediction of no difference in enjoyment for the low-self-esteem group was that while success would be inconsistent with their self-evaluation of their ability, it might be consistent with their cultural concept of enjoyment of tasks in which they succeed.

Predictions for the ratings of appropriateness of the first task for measuring intelligence were also based upon a self-consistency model. It was predicted that the pattern of these ratings would be similar to the enjoyment ratings pattern.

Similar but weaker relationships were predicted for ratings of appropriateness of Task 1 for studying problem-solving behavior. This prediction was based partly upon the assumption that *S*s will have much less ego involvement in ratings of appropriateness of Task 1 for studying problem-solving behavior than ratings of appropriateness of Task 1 for measuring intelligence. If this variable is of less psychological importance, distortion via self-consistency mechanisms should be smaller.

The predictions for the perseverance on the second task variable depend upon the stability of self-esteem after an experience of success or failure on Task 1. If self-esteem is not significantly altered by success or failure on Task 1, it would simply be predicted that high-self-esteem *S*s would persevere longer on Task 2 than low-self-esteem *S*s.

A strong argument could be made, however, that success or failure on Task 1 would have a significant effect upon perseverance on Task 2. One might expect that self-esteem would be increased or decreased with

TABLE 2

INTERCORRELATION MATRIX FOR THE DEPENDENT VARIABLES

Variable	E (Task 1) (5 = very enjoyable)	E (Task 2) (5 = very enjoyable)	M (Task 1) (5 = very appropriate)	M (Task 2) (5 = very appropriate)	P (Task 1) (5 = very appropriate)	P (Task 2) (5 = very appropriate)
E (Task 1)						
E (Task 2)	.48					
M (Task 1)	.21	.21				
M (Task 2)	.23	.19	.55			
P (Task 1)	−.06	−.03	.20	.04		
P (Task 2)	.00	.15	.12	.11	.63	
Perserverance (Task 2)	.15	.15	.11	.15	.05	.09

Note.—All intercorrelations of *S* ratings are based upon all *S*s ($N = 148$). Intercorrelations of *S* ratings and perseverance on Task 2 are based only upon *S*s who were not able to solve Task 2 ($N = 112$). The direction in which variables were scored is given below each variable. Pearson *rs* > .160 are significant at the $.05_2$ level and rs > .209 are significant at the $.01_2$ level for $df = 146$. Pearson *rs* > .165 are significant at the $.05_2$ level and *rs* > .232 are significant at the $.01_2$ level for $df = 110$. Abbreviations: E = enjoyment, M = appropriateness for measuring intelligence, P = appropriateness for studying problem-solving behavior.

success or failure on Task 1 and, therefore, the expectations for success or failure on Task 2 would also be altered somewhat.

The predicted results for the perseverance on the Task-2 variable allowing for the possibility of self-esteem being altered by an experience of success or failure on Task 1 are shown below. (SE = self-esteem.)

Perseverance on the Second Task

Low SE - Failure ≤ Low SE - Success <
High SE - Failure ≤ High SE - Success

This conjecture is based upon the concept that high-self-esteem *S*s who fail may persevere just as long as high-self-esteem *S*s who succeed on Task 1 in order to maintain their high self-esteem.

Predictions for ratings of enjoyment of the second task, appropriateness of the second task for measuring intelligence, and appropriateness of the second task for studying problem-solving behavior also depend upon the stability of self-esteem. For this reason, relationships between self-esteem, success, and failure on Tasks 1 and 2 and these three rating variables are bound to be quite complex. Further, from pretest data, the percentage of *S*s succeeding on Task 2 was expected to be too low for meaningful analyses of these data in terms of the self-consistency model, and no specific predictions are presented here for these variables.

The experimental results can logically be divided into two components—results obtained from *S*s who were *not* able to solve the second task and results from *S*s who *were* able to solve that task. The predictions for perseverance on the second task must be based upon the results from the first group of *S*s. The data from the second group of *S*s, however, are of considerable interest. A comparison of *S*s in both groups gave information as to what type of person not only persevered on the second task but also solved it. An analysis of these data was made to see if these patterns fit into a self-consistency framework.

RESULTS

Variable Intercorrelations

Table 1 gives the intercorrelation matrix for the independent variables in the present experiment: task success, self-assurance, self-confidence, personal adequacy, and set for all *S*s.

The magnitude of the self-esteem variable intercorrelations (.30, .26, .25) does not provide strong support for the assertion that these variables closely reflect the same personality attribute.

The sex variable is significantly related at the .05 level (two-tailed test) to the other five independent variables and were, therefore, included in all other inferential analyses.

Table 2 gives the intercorrelation matrix for the dependent variables, enjoyment of Tasks 1 and 2, appropriateness of Tasks 1 and 2 for measuring intelligence, appropriateness of Tasks 1 and 2 for studying problem-solving behavior, and perseverance on Task 2.

All correlations between *S* ratings of Tasks 1 and 2 and perseverance on Task 2 are nonsignificant at the .05 level. The lack of significant relationships suggests that task perseverance cannot be taken as a measure of task enjoyment or liking as indicated by a rating.

TABLE 3

ANALYSIS OF THE EFFECTS OF SELF-ESTEEM, SUCCESS OR FAILURE ON TASK 1, SEX, AND INTERACTIONS UPON RATINGS OF ENJOYMENT OF TASK 1 FOR ALL *S*s

Variable	$(R_L^2 - R_S^2)$	$(1 - R_L^2)$	*F*	*df*	*p*
Analysis 1—SA					
SA	.00	.92	.31	1,144	.58
S-F	.05	.92	8.44	1,144	.01
Sex	.01	.92	1.56	1,144	.20
SA × S-F	.00	.92	.00	1,143	1.00
Analysis 2—SC					
SC	.01	.91	1.27	1,144	.26
S-F	.06	.91	8.89	1,144	.00
Sex	.02	.91	2.54	1,144	.11
SC × S-F	.01	.91	1.41	1,143	.24
Analysis 3—PA					
PA	.00	.92	.16	1,144	.69
S-F	.05	.92	8.44	1,144	.01
Sex	.01	.92	1.56	1,144	.22
PA × S-F	.00	.92	.00	1,143	1.00

Note.—Abbreviations: SA = self-assurance, S-F = success-failure, SC = self-confidence, PA = personal adequacy. $N = 148$.

Inferential Tests of Main and Interaction Effects

Least-squares analyses, obtained through regression methods as proposed by Cohen (1968) and Overall and Spiegel (1969), have been used in the treatment of the present experimental data. The multiple-regression technique used here provides an ANOVA analogue solution that can easily handle unequal cell frequencies and does not require artificial dichotomy of the self-esteem variables.[2]

Results of analyses of the effects of self-esteem (Ghiselli [undated] self-assurance [SA], Nelson [1968] self-confidence [SC], and Nelson [1968] personal adequacy [PA]), success or failure on Task 1 (S-F, Task 1), sex, and interactions upon ratings of enjoyment of Task 1 are given in Table 3.

The variation of *F* values for the success-failure and sex variables in the three analyses reflects the differing correlations between these variables and the self-esteem variables. The success-failure variable is significant at the .005 level with all other variables being significant at greater than the .10 level. The predicted significant Success-Failure × Self-Esteem interaction was not obtained in any of the three analyses.

In Table 4 the self-esteem variables have been dichotomized near the median points so that means and *t* tests could be computed and visual inspection made of the ratings of enjoyment of Task 1 data.

It should be noted that larger *t* values (though not significantly larger) were obtained for the comparison of success and failure conditions for low-self-esteem *S*s than for the same comparison for high-self-esteem *S*s. These trends are not only nonsupportive but are contrary to the predictions.

Analyses of results for ratings of appropriateness of Task 1 for measuring intelligence and studying problem-solving behavior yielded, for the most part, nonsignificant results.

The results of the analyses of the effects of self-esteem (SA, SC, PA), success or failure on Task 1, sex, and interactions upon perseverance on Task 2 were nonsignificant.

The predicted results were not obtained in the three analyses as the multiplicative effects yielded minimal *p* values. The success-failure effect was significant at the .10 level for all three analyses.

The self-esteem variables were dichotomized near median points so that means and *t* values could be computed (in seconds) on Task-2 data. In general, success *S*s on Task 1 persevered longer on Task 2 than failure *S*s. Low-

---

[2] The present design (Method 2, or the experimental design analysis in Overall and Spiegel's [1969] classification of least-squares procedures) involves estimation of main effects disregarding interactions and the estimation of interactions adjusted for main effects.

Multiplicative effects correspond to interactive effects in conventional ANOVA designs and were tested in the present design by multiplying any number of independent variables together for each *S* and including this new variable in the regression equation.

The *F* ratios reported here were obtained by applying the multiple R^2s obtained through standard regression techniques to the following equation:

$$F = \frac{(R_L^2 - R_S^2)/(L - S)}{(1 - R_L^2)/(N - L - 1)}$$

where R_L^2 and R_S^2 are the R^2s based upon the larger and smaller sets of independent variables, respectively, and *L* and *S* are the number of independent variables in the larger and smaller sets.

self-esteem *S*s who succeeded on Task 1 persevered longer (though not significantly longer) than high-self-esteem *S*s who succeeded on Task 1. Also, in general, low-self-esteem *S*s who failed on Task 1 persevered longer (though not significantly longer) on Task 2 than high-self-esteem *S*s who failed on Task 1. While none of the observed relationships is significant at the .10 level, even the trends noted are not consistent with the predictions.

As suggested earlier, the number of *S*s who succeeded in their attempts to solve the second puzzle was too small to allow for a full analysis of results in relation to success or failure on Tasks 1 and 2 and the self-esteem variables.

Since the self-esteem and Self-Esteem × Success-Failure interaction effects in all analyses involving ratings of Task 1 and perseverance on Task 2 were nonsignificant, these effects were ignored in analyzing ratings for Task 2. This procedure made it possible to compare the effects of various combinations of success or failure on Task 1 and 2 upon ratings of Task 2. Table 5 gives the means and standard deviations of ratings of enjoyment of Task 2, appropriateness of Task 2 for measuring intelligence, and appropriateness of Task 2 for studying problem-solving behavior in relation to various combinations of success or failure on Tasks 1 and 2.

TABLE 4

MEANS, STANDARD DEVIATIONS, AND *N*S FOR RATINGS OF ENJOYMENT OF TASK 1 IN RELATION TO SELF-ASSURANCE, SELF-CONFIDENCE, PERSONAL ADEQUACY, AND SUCCESS OR FAILURE ON TASK 1

Variable	Failure[a]			Success[b]			t
	$\bar{X}$	*SD*	*n*	$\bar{X}$	*SD*	*n*	
SA							
Low (0–25)	1.61	1.24	48	2.32	1.07	28	2.57*
High (26 up)	1.69	1.30	42	2.30	1.21	30	2.07*
SC							
Low (0–11)	1.53	1.30	53	2.65	1.11	34	4.18**
High (12 up)	1.81	1.20	37	1.83	1.03	24	*ns*
PA							
Low (0–14)	1.53	1.20	43	2.23	.88	28	*ns*
High (15 up)	1.77	1.33	43	2.39	1.37	28	*ns*

Note.—Abbreviations: SA = self-assurance, SC = self-confidence, and PA = personal adequacy; 5 = very enjoyable, 1 = very unenjoyable.

[a] t values for comparison of means for SA, SC, and PA were *ns*.

[b] t values for comparison of means for SA and PA were *ns*; for SC, $t = 2.816$.

* $p < .05_2$.

** $p < .01_2$.

TABLE 5

MEANS AND STANDARD DEVIATIONS FOR RATINGS OF ENJOYMENT OF TASK 2, APPROPRIATENESS OF TASK 2 FOR MEASURING INTELLIGENCE, AND APPROPRIATENESS OF TASK 2 FOR STUDYING PROBLEM-SOLVING BEHAVIOR IN RELATION TO SUCCESS OR FAILURE ON TASKS 1 AND 2

Measure	F_1F_2[a]	S_1F_2[b]	F_1S_2[c]	S_1S_2[d]
Enjoyment of Task 2				
$\bar{X}$	1.85	2.53	2.19	2.95
SD	1.33	1.41	1.11	1.13
Appropriateness of Task 2 for measuring intelligence				
$\bar{X}$	2.52	2.76	3.20	3.09
SD	1.19	.94	.98	.97
Appropriateness of Task 2 for studying problem-solving behavior				
$\bar{X}$	2.35	2.51	2.22	2.71
SD	1.21	1.03	1.22	1.08

Note.—F_1 = failure on Task 1; F_2 = failure on Task 2; S_1 = success on Task 1; S_2 = success on Task 2.

[a] $N = 75$.

[b] $N = 15$.

[c] $N = 37$.

[d] $N = 21$.

The mean differences between these four groups for the enjoyment of the Task-2 variable is significant at the .02 level ($F = 3.50$, $df = 3$, 144). Analyses of mean differences for the two other Task-2 rating variables yielded nonsignificant results (at the .05 level).

Comparison of Subjects Who Succeeded or Failed on Task 2

Only about 30% of all *S*s were able to solve the second puzzle when given as much time to work on it as they wished. Comparisons of success and failure *S*s differed significantly on only two of the seven major independent variables. It was found that a higher proportion of male *S*s successfully solved the second puzzle than female *S*s, and a higher propor-

TABLE 6

MEAN SATISFACTION WITH PERFORMANCE RATINGS IN RELATION TO INITIAL EXPECTANCY OF SUCCESS AND SUCCESS OR FAILURE ON THE TASKS

Initial expectancy	Failure		Success	
	Rating	*N*	Rating	*N*
Low	2.6	46	5.4	39
High	1.3	37	4.5	48

Note.—This table presents pooled data from several tables in the Feather (1969) study on a rating scale of 5 = very satisfied, 1 = very dissatisfied.

tion of those *S*s who solved the first puzzle were able to solve the second puzzle.

DISCUSSION

The results predicted for the ratings of enjoyment of Task 1 were not obtained in the present experiment. Task enjoyment was significantly related to task success, with low-self-esteem *S*s enjoying success as much or more than high-self-esteem *S*s. Self-esteem as measured by the self-assurance, self-confidence, and personal adequacy scales was minimally related to task enjoyment as a main effect or as a moderator effect.

The failure to support Korman's (1968) results is difficult to explain. The dependent variables and experimental instructions were very similar in both experiments. While the tasks employed were not the same, the self-consistency model would predict similar results for this situation. Both types of tasks are used in intelligence or aptitude batteries (the *S*s were told this in both studies) and the mean rating given to the appropriateness of Task 1 for measuring intelligence suggests that *S*s believed the task to be at least moderately well suited for measuring intelligence.

A recent study by Feather (1969) yielded results that are supportive of the findings in the present study. In the Feather experiment, measures of self-esteem and task-specific expectations of success were obtained prior to attempts by *S*s to solve anagram-type tasks. Feather predicted that self-esteem would be significantly related to task-specific expectancies and that low-self-esteem-expectancy *S*s who succeeded on these tasks would rate satisfaction with their performance on the anagrams task higher than high-self-esteem-expectancy *S*s who succeeded on the tasks. Feather also predicted that high-self-esteem-expectancy *S*s who failed on the tasks would rate satisfaction with their performance lower than low-self-esteem-expectancy *S*s who failed on the tasks.

Feather obtained supportive results for his hypothesis only when he used the initial task-specific expectancy scores, all self-esteem measures yielding nonsignificant results. By pooling Feather's (1969) results, it is possible to present his data so that it can be compared with the results obtained in the present experiment and the Korman (1968) experiment. Table 6 presents mean satisfaction ratings in relation to initial expectancy and success or failure on the tasks for the Feather study.

Visual inspection of the data suggest that success *S*s, regardless of initial expectancy, rated satisfaction with their performance higher than failure *S*s. These results are also contrary to balance model predictions but support the findings of the present study.

It should be noted that the satisfaction ratings in the Feather (1969) experiment were different in a subtle way from the satisfaction ratings in the present experiment and in Korman's (1968) experiment. In the Feather experiment, *S*s were asked to rate satisfaction with their performance, whereas in the present and Korman experiments, *S*s were asked to rate their enjoyment or liking of the tasks. One can conceive of a high-self-esteem *S* rating satisfaction with his performance low and enjoyment of the task considerably higher. The ratings in the Feather study may be influenced by a perfectionist effect.

The Pearson *r*s between initial confidence and the three measures of self-esteem (.17, .10, and −.16) were nonsignificant at the .05 level. Feather's results suggest that self-esteem as measured by available self-rating or evaluation tests may be rather poor measures of initial expectancies and should not be interpreted in this way.

The experimental results obtained in the present experiment are consistent with Locke's (1965) conception of the relationship between task success and task satisfaction and are also consistent with a naïve conception of the relationship between these two variables.

Results obtained for ratings of Task 2 in relation to success-failure on Tasks 1 and 2 are generally consistent with results obtained for ratings on Task 1. The *S*s experiencing success on Tasks 1 and 2 rated enjoyment of Task 2 higher than *S*s who experienced mixed success on Tasks 1 and 2, who in turn rated enjoyment of Task 2 higher than *S*s who failed on both Tasks 1 and 2.

The pattern obtained for the perseverance data was similar, though the relationships were weaker, to the patterns obtained for the enjoyment ratings for both tasks. The consistency of the contrahypothesized results in the present experiment suggest that some essential differences do exist between the experimental designs mentioned in this discussion.

If experimental tasks and dependent measure differences were responsible for the cross-experiment differences, then the balance or self-consistency model postulated by Korman (1968) is a great deal more complex than originally conceived, or it is situation bound.

It has also been suggested in this discussion that functional differences may exist between general measures of self-esteem (e.g., Ghiselli's self-assurance scale and Feather's [1969] measure of self-esteem) and task-specific measures of expectations of success. One might expect differences in the manner in which various tasks would interact with chronic self-esteem in this case.

The fact that success and failure were not induced or manipulated in the present study but were self-evident may also be a major difference in the two studies.

Analyses of differences on seven independent variables between *S*s who succeeded or failed on Task 2 revealed that the sex and success-failure on Task 1 variables were the only variables for which significant differences were noted between the two groups.

SUMMARY

In general, the findings of this study are not consistent with predictions based upon Korman's (1968) self-consistency model. The results obtained in the present study and the Feather (1969) study suggest that task success is significantly related to task enjoyment for low-self-esteem *S*s *as well as* for high-self-esteem *S*s. These findings are also consistent with results reported by Locke (1966) indicating a positive relationship between task success and task liking.

REFERENCES

ARGYRIS, C. *Integrating the individual with the organization*. New York: Wiley, 1964.

COHEN, J. Multiple regression as a general data-analytic system. *Psychological Bulletin,* 1968, **70**, 426–443.

DEUTSCH, M., & GERARD, H. A study of normative and informational social influence upon individual judgment. *Journal of Abnormal and Social Psychology,* 1955, **51**, 629–636.

DUNNETTE, M. D., CAMPBELL, J. P., & HAKEL, M. Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Organizational Behavior and Human Performance,* 1967, **2**, 143–174.

FEATHER, N. T. Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology,* 1969, **13**, 129–144.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957.

GAVIN, J. F. Ability, effort and role perceptions as antecedents of job performance. Unpublished doctoral dissertation, New York University, 1969.

KORMAN, A. K. Task success, task popularity, and self-esteem as influences on task liking. *Journal of Applied Psychology,* 1968, **52**, 484–490.

LECKY, P. *Self-consistency: A theory of personality.* New York: Island Press, 1945.

LOCKE, E. A. The relationship of task success to task liking. *Journal of Applied Psychology,* 1965, **49**, 379–385.

LOCKE, E. A. The relationship of task success to task liking: A replication. *Psychological Reports,* 1966, **18**, 552–554.

NELSON, R. *The Trait Evaluation Index: Examiner's manual.* New Rochelle, N. Y.: Martin C. Bruce, 1968.

OVERALL, J. E., & SPIEGEL, D. K. Concerning least-squares analysis of experimental data. *Psychological Bulletin,* 1969, **72**, 311–322.

SCHEIN, V. E. The relationship between accuracy of subordinate's perception of his supervisor and job performance. Unpublished doctoral dissertation, New York University, 1969.

VROOM, V. *Work and motivation.* New York: Wiley, 1964.

WEITZ, J. Tiny theories. *American Psychologist,* 1967, **22**, 157.

(Received June 22, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 422–432

LEADERSHIP STYLE, HIERARCHICAL INFLUENCE, AND THE SATISFACTION OF SUBORDINATE ROLE EXPECTATIONS:

A TEST OF LIKERT'S INFLUENCE PROPOSITION

ROBERT J. HOUSE[1]
Bernard M. Baruch College
City University of New York

ALAN C. FILLEY
University of Wisconsin

AND DAMODAR N. GUJARATI
Bernard M. Baruch College, City University of New York

This article deals with the moderating effect of upward hierarchial influence on the relationship between supportive leader behavior and satisfaction of subordinate expectations and with the effect of four dimensions of leadership style. Contrary to previous theory and research, the findings reveal little interaction between hierarchical influence and consideration. Leader consideration was found to have strong positive relationship to satisfaction in 11 out of 16 measures of subordinate role satisfaction. Contrary to previous research, initiating structure was found to be positively related to 10 out of 16 measures of subordinate role satisfaction. Leader technical competence was found to be significantly related to only 2 out of 16 role satisfaction measures, and leader decisiveness was found to have a rather pervasive effect on role satisfaction, bearing a significant positive relationship to 13 out of 16 satisfaction measures. The theoretical implications of the findings as related to Likert's linking-pin theory and previous studies on leader behavior are briefly discussed.

The present article explores the relationship between subordinate perceptions of the behavior of their immediate superiors and the satisfaction of subordinate role expectations. Subordinate expectations, to the extent that they are fulfilled, represent degrees of satisfaction with the job and the organizational environment.

The theoretical significance of the research reported in this article lies in the tests of hypotheses concerning the effects of several leadership behaviors that revealed findings contrary to previous research and theory. Of particular current interest is the test of an hypothesis concerning the effect of leadership influence as a moderating variable on which the relationship between leadership behavior and satisfaction with subordinate role expectations is contingent.

According to Likert (1961, 1967) upward hierarchical influence permits subordinate managers to obtain the assistance and support of their superiors and thereby provides for linkage with higher strata and status levels in the organization. In contrast with a downward influence system from superior to subordinate, the upward influence of the superior is seen as a measure of his ability to control resources for the subordinate work group and to represent the work group in its dealings with the broader organization system. Thus, upward hierarchical influence is seen by Likert as a necessary though not a sufficient condition affecting group satisfaction and productivity. For example, Likert (1961) asserts that the leader must not only have upward hierarchical influence with his immediate superior, but must also exhibit a supportive subordinate-centered leadership style. The supportive style is seen as building higher expectations on the part of subordinates. The satisfaction of these expectations is contingent on the ability of the leader to represent the group in the broader organization system or with higher levels in a manner that benefits the work group.

Thus, it would appear that leader behavior affects employee satisfaction and commitment to organization goals and that upward influence by the leader is primarily an intervening or moderating variable.

---

[1] Requests for reprints should be sent to Robert J. House, Bernard M. Baruch College, Baruch Graduate Center, City University of New York, 257 Park Avenue South, New York, New York 10010.

Dimensions of Leader Behavior Studied

Of the many dimensions of leader behavior in the research literature that are found to be related to satisfaction and performance in a work group, five would seem to be of paramount importance. These are initiating structure, consideration, decisiveness, technical competence, and hierarchical influence.

Initiating structure and consideration. These two dimensions of leader behavior describe the degree to which the leader establishes psychological structure for subordinates and creates an environment of psychological support, warmth, and friendliness. In reporting the early development of these dimensions, Halpin and Winer (1957) describe initiating structure behaviors as those which indicate the leader organizes and defines the relationships between himself and his subordinates; defines the role that he expects each subordinate to assume; and endeavors to establish well-defined patterns of organization, channels of communication, and ways of getting the job done. They describe consideration in terms of behavior indicative of friendship, mutual trust, respect, and warmth in the relationship between the leader and his subordinates.

Recent reviews of the evidence from a large number of studies indicate that leaders who initiate psychological structure for subordinates are generally rated high by superiors and frequently have higher producing work groups than leaders who are low on initiating structure, and that considerate leaders generally have more satisfied employees (see e.g., Fleishman, 1969). However, several studies have shown that structure is frequently resented by subordinates and is a source of dissatisfaction, turnover, and grievance (Fleishman & Harris, 1962; Korman, 1966; Filley & House, 1969).

Technical competence and decisiveness. Leader decisiveness refers to the degree to which the leader is willing to make decisions or take decisive action. Technical competence refers to the degree to which the leader is perceived as capable of providing advice on technical or specialized problems and capable of anticipating job-related details prior to assigning tasks. While these two dimensions of leader behavior have not received the attention that the first two have, there is evidence of their importance as correlates of leadership effectiveness. The studies by Comrey, Pfiffner, and High (1954) have indicated that supervisors of forest workers, government employees, and industrial production workers show significant positive relationships between these measures and measures of group productivity and leader effectiveness.

Hierarchical influence. Research reported by Pelz (1952) and Wager (1965) suggests the importance of upward influence by the leader as a variable on which the relationship between supportive leader behavior and the satisfaction of subordinate role expectations is contingent. In his study, Pelz found that influence served to moderate the effect of the supervisor's "social closeness" to employees. When supervisors had high influence and also engaged in supportive leader behavior by "taking sides with subordinates" rather than "taking sides with management," employee satisfaction was significantly greater than when supervisors with low influence engaged in the same kind of supportive behavior.

Similarly, in his study of the relationship between supportive behavior and influence in one firm, Wager (1965) found highly supportive behavior to be positively correlated with hierarchical influence on the part of first-level supervisors. He also found that influence had a moderating effect on the supportive behavior of leaders similar to the effect found by Pelz (1952) but that the magnitude of the moderating effect of influence varied positively with organizational status of the respondent. This finding led him to conclude that when employees are performing work that permits them to be independent of their immediate superiors, the influence of the superior is less relevant to the fulfillment of subordinate expectations. Here it is assumed that when employees enjoy high organizational status, they are more likely to have occupational mobility, to be more autonomous, and to be more self-sufficient in their work setting.

Hypotheses

The previous research and theory suggest the following hypotheses to which the present study is addressed:

1. The relationship between the leader's consideration and subordinate role satisfactions will be moderated by the amount of influence imputed to the leader by the subordinates. Specifically, where the superior is perceived as having high hierarchical influence, the relationship between leader consideration and subordinate role satisfaction will be positive and significantly greater than when the superior is perceived to have low hierarchical influence.
2. Leader consideration will be positively related to the amount of influence imputed to the leader by his subordinates.
3. Leader consideration will be positively related to the satisfaction of subordinate role expectations.
4. Leader structure will be negatively related to the satisfaction of subordinate role expectations.
5. Leader technical competence will be positively related to the satisfaction of subordinate role expectations.
6. Leader decisiveness will be positively related to the satisfaction of subordinate role expectations.

In addition to tests of the above hypotheses, exploratory tests were also conducted concerning the moderating effect of influence on the relationship between: (*a*) the leader's technical competence, initiating structure, and decisiveness; and (*b*) subordinate role expectations. These tests are considered exploratory because neither research nor theory suggest specific hypotheses concerning these interactions.

METHOD

Research Design

The data for the study were collected from the research, development, and design departments of a large petroleum refining company and a large business machine manufacturing company. Samples, which represented 50% of the populations studied, were selected randomly and stratified by organizational level.

Questionnaires were administered in group settings of approximately 20–35 persons in the company cafeterias during working hours. Respondents were assured that all responses would be anonymous and confidential. Company A completed 126 questionnaires, and Company B completed 242. Of these, 84 in Company A and 159 in Company B contained complete information. To facilitate analysis, only questionnaires with complete data were used in this study. Comparisons of the scale means and standard deviations of the questionnaries used here and the questionnaires with some missing data revealed no significant differences between those who failed to complete some items and those who provided full information.

Measures Employed

Satisfaction of employee role expectations. Two questionnaires, both developed by the Personnel Research Board of the Ohio State University, were used. The first was the Job Description questionnaire that contains 12 questions scaled 1–5. This questionnaire essentially measures the employees' attitude toward the company and its management (Stogdill, 1965b).

The remaining seven measures of satisfaction with role expectations are contained in a Job Expectation questionnaire. Employee satisfaction of role expectations is measured with respect to the following classes of role expectations: (*a*) work, (*b*) freedom, (*c*) respondent's job as compared with jobs of others, (*d*) pay, (*e*) advancement, (*f*) family attitudes toward the respondent's job, and (*g*) job security.

Respondents are asked to indicate the degree to which each characteristic of the job satisfies his expectations. Each class of expectation is measured by scores of seven questions.

Leader consideration and initiating structure. A form of the Ohio State Leader Behavior Description Questionnaire (LBDQ) described by Halpin (1957) was used to measure these two dimensions of leader behavior.

Leader decisiveness, technical competence, and hierarchical influence. The scales used to measure these behaviors were factorially developed by Comrey, Pfiffner, and High (1954). These scales consist of four, five, and four items, respectively, and, like the LBDQ, are completed on a 5-point multiple-choice response mode indicating the frequency with which the leader engages in the behavior described by the item stem.

Following is one sample, item for each scale:

Decisiveness—He acts cautiously in order to avoid sticking his neck out. (1) frequently; (2) occasionally; (3) rarely; (4) almost never; (5) never.

Technical Competence—When a tough job comes up, he has the technical "know how" to get it done. (1) rarely; (2) occasionally; (3) usually; (4) almost always; (5) always.

Hierarchical Influence—He has requested from his superiors certain benefits for his workers which have not been approved. (1) frequently; (2) occasionally; (3) rarely; (4) almost never; (5) never.

It is important to point out that the influence scale (and the studies by Pelz [1952] and Wager [1965]) deals with the upward influences of the leader *only* with respect to matters concerning the personnel management of his subordinates and *does not* measure influence with respect to such decisions as resource allocation, organization changes, objective setting, or policy formulation.

Stogdill (1965) reports that reliabilities of measures of initiating structure, consideration, and hierarchical influence consistently fall in the range of .80 and above for several widely different populations including United States Senators, company presidents, middle managers, military officers, and state police officers.

These data reported by Stogdill attest to the internal consistency of the leader behavior scales. However, they do not clearly indicate that the scales measure what they are purported to measure. However, Halpin (1957) has shown the LBDQ to have concurrent criterion validity and a recent experiment reported by Stogdill (1969), conducted under well-controlled laboratory conditions, resulted in rather clear experimental criterion validation of the Ohio State scales. The study by Comrey, Pfiffner, and High (1954) suggests construct validity of the Leader Decisiveness, Influence, and Technical Competence scales. In an extensive series of field investigations they found these three measures to have factorial independence and to have significant correlations with objective measures of work unit performance such as scrap and output as well as subjective ratings of managerial performance by superiors.

Method of Analysis

A modified analysis of covariance using dummy variables in a regression was employed. Respondents in each company were trichotomized on the influence variable with the upper and lower thirds representing high- and low-influence leaders.[2] The high- and low-influence conditions are treated as dummy variables for regression analysis. That is, the variables are coded so that each enters the model separately to yield a regression coefficient for the relationship between leader consideration and satisfaction with role expectations under each of the influence conditions.

To test the hypothesis that the relation between consideration and satisfaction with role expectations will be moderated by the amount of leader influence, the following regression model was specified and tested.[3]

$$S = a_0 + a_1 D_1 + a_2 D_2 + a_3 Y + a_4 YI_H + a_5 YI_L + e \quad [1]$$

where: S = scalar value of satisfaction; $D_1 = 1$ when influence is high, 0 otherwise; $D_2 = 1$ when influence is low, 0 otherwise; Y = scalar value of independent variable; YI_H = scalar value of independent variable under high influence; YI_L = scalar value of independent variable under low influence; e = stochastic error term assumed to be normally and independently distributed with zero mean and constant variance; a_0, a_1, and a_2 = intercept coefficients; a_3, a_4, and a_5 = slope coefficients.

The Parameters a_0–a_5 can be interpreted as follows: a_0 = value of the intercept term under medium influence, $a_1 + a_0$ = value of the intercept term under high influence, $a_2 + a_0$ = value of the intercept term under low influence, a_3 = value of the slope coefficient under medium influence, $a_4 + a_3$ = value of the slope coefficient under high influence, $a_5 + a_3$ = value of the slope coefficient under low influence.

From the structure of Equation 1 it is obvious that Parameters a_1 and a_2 are the differential intercept coefficients, and the Parameters a_4 and a_5, the differential slope coefficients. Because of the form of our hypothesis, we are primarily concerned with the slope coefficients under the different amounts of influence. If a_4 and a_5 are significant, the relation between S and Y under various levels of influence will be significantly different. Of course, all the differential coefficients may not be statistically significant. If a_4 or a_5 have no effect, a_3 will yield the linear value of the full range of influence. If a_4 or a_5 has a significant effect, a_3 will represent only the effect of medium influence. According to our hypothesis, we would expect both consideration (a_3) and consideration in interaction with high influence ($a_3 + a_4$) to have significant effect.

The above model was tested with regression analysis and the same procedure was computed four times for each company substituting the other three dimensions of leadership style for consideration, that is, letting the Y variable take on the value of leader initiating structure, decisiveness, and competence. In total, 64 regressions were run (one regression for each of eight independent variables, for each of two companies, for each of four leadership style dimensions, or, eight satisfaction measures × two companies × four leadership style measures).

To test Hypotheses 2–6, product-moment correlations were computed for the relationships between the independent and dependent variables and a one-tailed test of significance was applied.

RESULTS

Hypothesis 1

The first hypothesis stated that influence would have a positive moderating effect on the relationship between leader consideration and subordinate satisfaction. Table 1 presents the significant regression coefficients.

For Company A, consideration has no effect on any satisfaction variables except family attitudes under any degree of influence. For satisfaction with family attitudes toward the job, there is a positive effect on slope, which is the same under high and low influence.

---

[2] Trichotomization results in unequal cell frequencies. The reason for the regression formulation is that:

> Unequal cell frequencies preclude the simple partitioning of the sum of the squares. The computation needed to generate the test statistics now involve matrix inversion in least-squares regression. However, from the point of view of hypothesis testing the two techniques (the use of dummy variables in regression analysis and analysis of variance or covariance) are equivalent [Galbraith & Cummings, 1967, p. 251; see also Goldberger, 1964, p. 230].

[3] This procedure also has the advantage of eliminating statistical bias due to colinearity among the predictor variables since no two predictors with full scalar values are included in the regression. Only the full value of Y and the partitioned thirds of influence are included.

TABLE 1

Significant Coefficients for Consideration and Consideration in Interaction with Influence as a Predictor of Satisfaction

Parameter	Company A: Satisfaction with family attitudes	Company B: Satisfaction with: Job	Work	Job freedom	Family attitudes	Security
a_1	−6.54*	—	5.15*	−6.36*	−5.19**	−7.17*
t	−1.95		2.30	2.13	−1.9	−2.43
a_2	−6.97	—	—	—	—	—
t	−2.02					
a_3	—	.12*	—	.10*	—	—
t		2.30		2.19		
a_4	.17*	—	− .16*	.16*	—	.15*
t	2.14		−2.15	2.28		2.16
a_5	.18*	—	—	—	—	—
t	2.16					

Note.—Only columns with significant values have been included here.
* $p = .05$.
** $p = .01$.

It would be a fair generalization to say that for Company A, consideration under any kind of influence has very little effect on satisfaction.

For Company B, the regression results show a different pattern. The slope effect (a_3) is positive under medium influence for satisfaction with work and job freedom; it is positive and higher under high influence ($a_3 + a_4$) for satisfaction with job freedom and security; it is negative under high influence for satisfaction with pay. The slope coefficient under low influence generally does not differ significantly from that under medium influence.

TABLE 2

Intercorrelations between Leadership Style Dimensions

Dimension	1	2	3	4	5
Company A[a]					
1. Structure	—	−.00	.21*	.08	.32**
2. Consideration		—	.38**	.18	.33**
3. Decisiveness			—	.09	.45**
4. Influence				—	.18
5. Competence					—
Company B[b]					
1. Structure	—	.54****	.41****	.41****	.14
2. Consideration		—	.56****	.30****	.21****
3. Decisiveness			—	.36****	.30****
4. Influence				—	.20****
5. Competence					—

[a] $N = 84$.
[b] $N = 158$.
* If $r = .18$, $p = .05$.
** If $r = .26$, $p = .01$.
*** If $r = .16$, $p = .05$.
**** If $r = .21$, $p = .01$.

So far as the hypothesis is concerned then, the empirical results show that influence has practically no effect on the relationship between satisfaction and consideration for Company A. For Company B, there is a pronounced effect for satisfaction of role expectations concerning freedom on the job and job security. Influence moderates the relationship between consideration and satisfaction in the hypothesized direction.

Hypothesis 2

The second hypothesis stated that leader consideration would be positively related to the amount of influence imputed to the leader by subordinates. Table 2 shows the intercorrelations between the five leadership style dimensions. As indicated, the correlation between consideration and influence is significant in Company B only. It approximates signifi-

TABLE 3

CORRELATIONS BETWEEN FIVE DIMENSIONS OF LEADERSHIP STYLE AND EIGHT DIMENSIONS OF SUBORDINATE SATISFACTION WITH ROLE EXPECTATIONS

Expectation	Leadership style									
	Consideration		Structure		Decisiveness		Influence		Competence	
	Company A	Company B	Company A	Company B	Company A	Company B	Company A	Company B	Company A	Company B
Company satisfaction	.25[a]	.55[b]	.25[a]	.47[b]	.35[b]	.42[b]	.20	.37[b]	.39[b]	.13
Intrinsic job expectation	.31[b]	.32[b]	.30[b]	.25[b]	.37[b]	.31[b]	.24[b]	.12	.16	.12
Advancement expectation	.35[b]	.43[b]	.06	.28[b]	.41[b]	.39[b]	.24[b]	.13	.16	.12
Comparison of job to jobs of friends	.19	.34[b]	.20	.20[a]	.30[b]	.26[b]	.15	.06	.12	.06
Pay expectation	.16	.29[b]	−.08	.21[a]	.23[a]	.27[b]	.12	.23[b]	−.08	.04
Expectation of freedom on the job	.46[b]	.36[b]	.02	.12	.31[b]	.39[b]	.19	.02	.15	.16[a]
Expectation of family attitudes toward the job	.27[a]	.15	.33[b]	.21[b]	.09	.45[b]	.14	.09	.10	.05
Job security expectation	.18	−.11	.25[a]	−.01	.13	−.14	.18	−.05	.14	−.03

Note.—$N = 84$, in Company A; $N = 158$ in Company B.
[a] $t \leq .05$ if $r \geq .22$ when $N = 84$; and $\geq .16$ if $N = 158$.
[b] $t \leq .01$ if $r \geq .28$ when $N = 84$; and $\geq .21$ if $N = 158$.

cance in Company A ($r = .178$, $p = .06$). This finding corroborates Wager's (1965) finding that influence was positively correlated with supportive leader behavior. Thus, it appears that there is a general relationship between supportive leadership and leader influence, but the magnitude of the relationship is unstable and dependent on situational factors associated with the population studied.

Hypothesis 3

The third hypothesis stated that leader consideration would be positively related to employee satisfaction with role expectations. Table 3 shows the correlations between the eight measures of employee satisfaction and the five dimensions of leadership style. In this case, 15 of the 16 relationships between consideration and satisfaction are positive, and of these, 11 are statistically significant. Thus, consistent with prior findings, there appears to be rather strong support for the hypothesized general relationship.

Hypothesis 4

The fourth hypothesis stated that leader initiating structure would be negatively related to the satisfaction of employee role expectations. As indicated in Table 3, the findings are strikingly contrary to those hypothesized. Initiating structure is negatively correlated in only two cases, and there the coefficients are quite low and not significant. Contrary to our predictions, in 10 of the 16 correlations, they were significantly positive and none were significantly negative.

Hypothesis 5

The fifth hypothesis stated that leader technical competence would be positively related to employee satisfaction. As shown in Table 3, technical competence is significantly related to the measures of satisfaction in only 2 of the 16 correlations.

Hypothesis 6

The sixth hypothesis stated that leader decisiveness would be positively related to measures of employee satisfaction. Table 3 shows that of the 16 possible relationships, 13 are significant in the expected direction, and none are significantly negative. Thus, there is surprisingly strong support for the hypothesis

TABLE 4

Significant Coefficients for Decisiveness and the Interaction of Decisiveness with Influence as a Predictor of Satisfaction for Company B

Parameter	Satisfaction with:					
	Company	Advancement	Pay	Job freedom	Family attitudes	Job security
a_1	—	—	—	−5.61*	−6.16**	−5.09*
t				−2.19	−2.56	−1.99
a_2	—	—	—	—	—	—
a_3	.614*	.533**	.21*	—	—	− .33**
t	2.09	3.71	2.17			−2.55
a_4	—	—	—	.39*	.43**	—
t				2.25	2.64	
a_5	—	—	—	—	—	.33*
t						2.14

Note.—Only columns with significant values have been included here.
* $p = .05$.
** $p = .01$.

that leader decisiveness is related to employee satisfaction.

Exploratory Analysis

In addition to tests of the above hypotheses, exploratory analyses among the dimensions of the leader behavior and of the moderating effect of influence were conducted. These analyses are viewed as exploratory rather than tests of hypotheses because there is neither previous research nor theory to suggest specific hypotheses to be tested.

Influence in interaction with other leader behavior dimensions. To explore further the moderating effects of influence, the regression procedure described above was repeated three times for each company, substituting the dimensions of initiating structure, decisiveness, and technical competence for Y in the regression, that is, for consideration. No significant coefficients were obtained in Company A for interactions between leader behavior and influence. None were obtained in Company B for technical competence or for the Technical Competence × Influence interaction. For Company B, however, several significant coefficients were obtained for initiating structure and decisiveness. The significant coefficients for Company B are presented in Tables 4 and 5.

These tables reveal that for Company B the influence condition under which significant interactions occur and the sign of the parameters are highly inconsistent. In Table 4, the high-influence condition is shown to significantly moderate the relationship between leader decisiveness and role satisfaction with job freedom and family attitudes in a positive direction. The low-influence condition moderates the relationship for satisfaction with job security, and the medium-influence condition moderates the relationship for satisfaction with company, advancement, and pay. The intercept values under the high-influence condition is negative for satisfaction with freedom, family attitudes, and job security. These findings suggest that low influence serves little or no purpose in enhancing the effects of leader decisiveness, that high influence may have dysfunctional consequences, and that there is a midrange in which influence is optimal for leadership purposes.

From Table 5, it can be seen that the high-influence condition has a negative moderating effect on the relationship between job satisfaction and initiating structure, and the intercept term for three satisfaction variables are higher under high influence.

Relationships among the dimensions of leadership style. Table 2 presents the correlation coefficients for the relationships among the measurements of leadership style. As can be seen from Table 2, Company A, initiating structure has a moderate positive relationship with technical competence and is relatively independent of the other measures of leader behavior. Consideration is moderately correlated with decisiveness and competence in Company A and slightly correlated with influence. In Company B, consideration is moderately correlated with decisiveness and influence and slightly correlated with competence. Technical competence is moderately correlated with structure, consideration, and decisiveness.

In Company B, excluding technical competence, all of the relationships among the independent variables are positive and moderately high. Technical competence has low positive correlations with all variables except decisiveness, which is positive and moderately high. The finding that the leadership style measures are differentially intercorrelated from company to company again suggests the importance of situational factors in leadership research.

Discussion

The above findings have several important implications for organization theory and suggest several questions for future investigation.

Influence as a Moderator of Leader Behavior

That influence was not an important moderator in one of the two research and development organizations reported in this study, a finding quite inconsistent with Likert's (1961) notion, is perhaps explained by Wager's (1965) research. Wager concluded that when employees are independent of their superiors, the influence of the superior is less relevant to the fulfillment of the employee's expectations. If the research and development personnel in Company A, like Wager's high occupational-status groups, have sufficient freedom and control over their environment so that they do not have to use their superior as a spokesman or as a source of resources, then our findings could be explained by Wager's interpretation.

A second interesting finding concerns the specific amount of influence found to have the most pervasive moderating effect. In Company B, the medium-influence condition more pervasively moderated the relationships between initiating structure, decisiveness, and satisfaction of subordinate role expectations. One interpretation is that the high-influence condition permits the superior to exercise so much control over subordinates that the relationship between superior and subordinate, *as viewed by the subordinate*, becomes highly delicate. Where supervisors are seen to have such high influence, it is likely that there will be greater status separation between them and their subordinates, and that such status differentiation will result in a restriction of upward information flow, less willingness on the part of subordinates to approach superiors, and less satisfaction with the social climate of the work unit.

TABLE 5

Significant Coefficients for Initiating Structure and the Interaction of Initiating Structure with Influence as a Predictor of Satisfaction for Company B

Parameter	Satisfaction with:			Job freedom
	Company	Job	Advancement	
a_1	12.2*	7.15*	—	6.07*
t	1.94	2.28		1.97
a_2	—	—	—	—
a_3	—	—	.14*	.11*
t			2.39	1.91
a_4	—	$-.17$*	—	—
t		-2.19		
a_5	—	—	—	—

Note.—Only columns with significant values have been included here.
* $p = .05$.

The above findings, and the speculative interpretations advanced here, suggest a need for reconsideration of the influence–linking-pin proposition.

Reconsideration of the Influence–Linking-Pin Proposition

If hierarchical influence is to enhance the functioning of the linking pin, it appears that the members of the group must be of marginal organizational status with little autonomy. Since the research from which Likert (1961) develops his proposition was conducted in hierarchical organizations, it is possible that the role of influence only functions in the manner Likert asserts for low-status organizational members who are moderately to highly dependent on superiors for satisfaction. It also appears that the optimum amount of influence relationships are statistically significant but not particularly strong. If it were found that in discussion groups operating outside the context of direct observation, these interrelationships failed to reach significance, it would suggest that the nature of the research situation itself is affecting the results attained in these earlier studies. Data collected as a part of the study described below were analyzed to examine this possibility. A portion of this study has been published elsewhere (Wilson, 1969).

METHOD

Procedure

Male undergraduates were recruited for part-time jobs. They were offered $6.00 to work in a group for one afternoon. Scheduling was accomplished over the telephone for work with the university's freshman orientation committee (a ficticious organization) at which time the secretary casually asked each applicant whether he knew any of the people already signed up for a particular session. If a potential *S* indicated that he was previously acquainted with any of the others, either he or his acquaintance was contacted for rescheduling with an appropriate excuse.

Twenty groups of five *S*s each met in a room with a one-way mirror and hidden microphones. Since the experiment was to be disguised as a job, certain techniques were used to make this seem more feasible. For example, the one-way mirror was covered with a fiberglass curtain that, while not obscuring the mirror completely, made it appear unlikely that anyone would be watching from the other side while it was drawn. However, it was in fact possible to observe *S*s through the drawn curtain.

All *S*s were told that they were working for the freshman orientation committee, which supposedly was exploring the feasibility of having freshmen discuss human-relation stories written by upperclassmen about common campus problems. Ten of the groups (observed) were observed by *E* who remained in the room during the group meeting. A second *O* remained behind the one-way mirror. The remaining 10 groups (nonobserved) were left alone in the room but were observed from behind a one-way mirror (covered by a see-through curtain). These *S*s were not told that they were being observed.[4]

At five equally spaced points (10-min. periods) throughout the discussions (2 hr. and 20 min.) *O* counted the number and content of communications initiated by each group member. The reliability coefficient between *O*s in counting total number of initiations for each *S* was .97. The members in each group were ranked from one to five according to how much they participated in the group discussion. The two members in each group with the highest rates of participation were considered high participators and the remaining three, low participators.

At the end of the session, each group member was asked to rank the others in the group as to quality of leadership. The sum of the ranks assigned to each individual by the other *S*s were taken as his leadership score. The leadership scores of the five persons in a group were then rank ordered from one to five. The top two ranks in each discussion group were placed in the high-leadership group, and the remaining three members were placed in the low-leadership group.

The *S*s were also given Bass' (1962) Orientation Inventory, which includes three mutually exclusive scales. Each person's score on the Self-Orientation scale was taken as a measure of dominance or ascendance. The results of earlier studies using the Self-Orientation scale seem to justify this choice.[5]

RESULTS

In order to determine the relationship between participation and self-orientation in the observed and nonobserved groups, the self-orientation scores of high and low participants were compared under both conditions. From Table 1 it is seen that while high participators tend to have higher self-orientation scores than low participators in both conditions, only in the observed conditions is this difference statistically significant.

To determine the relationship between leadership and self-orientation, the self-orientation scores of the high and low leadership ranks were compared in the observed and nonobserved conditions. It is seen in Table 1 that only in the observed conditions do those high in leadership have higher self-orientation scores. Also, only in the observed conditions

---

[4] In the study there were actually four conditions but only the observed-nonobserved manipulations were successful. For the full set of instructions, see Wilson (1967).

[5] For example, in a review of studies using the Orientation Inventory, Bass (1965) reports that the self-oriented *S* is "one who likes to assert himself, to compete, to stand out and be visible [p. 12]." Self-orientation has been positively correlated with aggressiveness-competitiveness (.23) on the Sixteen Personality Factor Questionnaire (Cattel, Saunders, & Stice, 1957) and on the Thematic Apperception Test, self-orientation correlated with the need for aggression (.25) and with the need for autonomy (.22). In small discussion groups, self-orientation correlated positively with the following peer ratings: "(*a*) concerned about himself and his own personal needs and feelings"; (*b*) "takes unjustified assumptions"; (*c*) dominates and imposes his will on the group (Bass, 1965)."

TABLE 1

MEAN SELF-ORIENTATION SCORES BY PARTICIPATION, LEADERSHIP, AND EXPERIMENTAL CONDITIONS

Score	Observed					Nonobserved				
	$\bar{X}$ (%)	*N*	*SD*	*t*	*p*	$\bar{X}$ (%)	*N*	*SD*	*t*	*p*
Participation										
High	31.10	20	5.36	3.28	.01	28.55	20	6.24	1.09	*ns*
Low	25.27	30	6.49			26.63	30	5.69		
Leadership[a]										
High	30.24	21	5.53	2.47	.02	27.05	20	6.90	.47	*ns*
Low	25.69	29	6.83			27.87	30	5.20		

Note.—Computation of the standard deviations and *t* are based on the arcsine transformation of the original figures. Two-way analyses of variance showed a participation main effect but no leadership effect. In both cases there was no observer effect, but the interaction effects were significant at approximately the .10 level.

[a] In one observed group, there was a tie for second place rank on leadership. The conservative resolution was to include both *Ss* in the high-leadership group. Thus, there were 21 *Ss* in high leadership in the observed condition.

is there a significant difference between the high- and low-leadership ranks.

The association between leadership and participation ranks was examined for *S*s in the observed and nonobserved groups. Table 2 shows the results of this analysis. In the observed groups there is a strong positive association between participation and leadership. In the nonobserved groups no such relationship exists.

Additional Findings

Some further findings are relevant at this point. Each communication act initiated by *S* was classified according to one of the following categories: (*a*) positive social-emotional (reinforcement, agreeing, joking); (*b*) negative social-emotional (disagreement, rejecting, antagonism); (*c*) task oriented (all interaction concerning the task); (*d*) task irrelevant (out-of-field interaction not classified as positive or negative social-emotional); (*e*) positive task morale (positive attitudes about the task or situation as an object); and (*f*) negative task morale (negative attitudes about the task or situation as an object). From this, the percentage of each *S*'s total communication acts falling into each of the six areas could be determined. This made it possible to determine whether there were any differences in communication content profiles between the high and low *S*s (on leadership, participation, and self-orientation) in the two experimental conditions.

The only communication category that differentiated between the high and low *S*s on any of the three variables was positive task morale. The results are seen in Table 3. In the observed condition, *S*s high on leadership and participation initiated higher rates of positive task morale. In addition, in the observed condition self-orientation was significantly correlated (Pearsonion *r*) with positive task morale (.37, $p < .01$), while in the nonobserved condition it was not (−.13).

The positive and negative task morale categories were included in the study to determine how the presence or absence of *O* affected *S*'s willingness to verbalize attitudes about any aspect of the situation (e.g., room, *O*) or the

TABLE 2

RELATIONSHIP BETWEEN LEADERSHIP AND PARTICIPATION IN OBSERVED AND NONOBSERVED GROUPS

Participation ranks	Observed leadership rank[a]			Nonobserved leadership rank[b]		
	High	Low	Total	High	Low	Total
High	17	3	20	10	10	20
Low	4	26	30	10	20	30
Total	21	29	50	20	30	50

[a] $\chi^2 = 25.30$, $p < .001$, $\phi^2 = .51$.
[b] $\chi^2 = 1.38$, $\phi^2 = .03$.

TABLE 3

POSITIVE TASK MORALE RATES BY PARTICIPATION LEADERSHIP AND EXPERIMENTAL CONDITIONS

Task morale rate	Observed					Nonobserved				
	$\bar{X}$ (%)	*N*	*SD*	*t*	*p*	$\bar{X}$ (%)	*N*	*SD*	*t*	*p*
Participation										
High	1.89	20	3.08	3.69	.001	1.20	20	4.18	1.91	*ns*
Low	0.71	30	3.62			0.54	30	3.55		
Leadership[a]										
High	2.12	21	2.47	5.26	.001	1.14	20	4.34	.94	*ns*
Low	0.64	29	3.53			0.63	30	3.77		

Note.—Computations of the standard deviation and *t* are based on the transformation of the original figures. The unit of analysis was the percentage of each *S*s total interaction categorized as positive task morale. Two-way analyses of variance showed participation and leadership main effects but no observation effects. The interaction effects were significant at approximately the .10 level in both cases.

[a] In one observed group there was a tie for second place rank on leadership. The conservative resolution was to include both *S*s in the high-leadership group. Thus, there are 21 *S*s in high leadership in the observed condition.

task.[6] Thus, those who ultimately were chosen as leaders in the observed group were likely to have said something pleasant or constructive about the situation. In the nonobserved condition there was a tendency for leadership to be assigned to those who had initiated higher rates of task-oriented interaction, but this did not quite reach statistical significance ($p < .10$).

The postsession interviews provide some cues about the perceptions of *S*s in the observed condition. When asked what they thought *O* was looking for, 60% of the observed *S*s stated in various ways that *O* was looking for patterns of leadership, dominance, influence, or differences in participation. The *S*s in the nonobserved condition gave no indication that they saw these issues as having any relevance to their being in the situation. It is likely that the salience of these issues in the observed condition affected the perception of *S*s so that they consciously or unconsciously used the same criteria in choosing leaders that they felt *O* was using.

DISCUSSION

This section, in offering some possible explanations for the differences reported above takes some cues from those who have analyzed the social psychology of the experiment (Orne, 1962; Riecken, 1958; Rosenberg, 1965). Thus, our interpretations are in terms of the meanings that the observed and nonobserved conditions held for *S*s and how these effected their behavior in the two situations. These are admittedly post hoc explanations but it is hoped that they provide some hypotheses for further research.

Following the postulations of Riecken (1958) and Orne (1962), it would be expected that with *O* in the room each group member is likely to be concerned with how he is being evaluated in the eyes of *O*. This concern, or what Rosenberg (1965) calls "evaluation anxiety," should affect various personality types in different ways. For those *S*s with high self-oriented needs, this situation would serve as a perfect arena for calling attention to himself and expressing his various self-centered needs (see Bass, 1965). This type of person would want to be highly visible and, in a leaderless discussion group, one must compete for visibility and status by participating highly (see Carter, Haythorn, Shriver, & Lanzetta, 1950).

---

[6] Such interaction was not classifiable as task-oriented communication because it had the task or situation as an object, rather than being directly involved in reaching the group goals (the writing of the stories). Examples of interaction that would be classified in these categories might be: (*a*) "I think this work is worthwhile" (positive task morale) or (*b*) "This whole job is a farce" (negative task morale). The reliability coefficient for this category was .48, which is rather low because little of this interaction was recorded. Therefore, only a few disagreements between *O*s contributed a high proportion of error.

It is interesting that both observers independently reported the impression that the high participators in the observed groups were more concerned with putting on a show for *O* than with actually accomplishing the task—traits that characterize the highly self-oriented person. Further evidence that the high participators, who were also high in self-orientation, were playing roles for *O* is seen in the fact that they initiated high rates of positive task morale interaction. This particular group of observed *S*s (high self-orientation) seems to be highly motivated to ingratiate themselves with *O*.

In a situation where evaluation-anxiety is high and interaction strained, group members should be gratified by those who are willing to carry out what Redl (1955) calls the "guilt-and-fear assuaging effect of the initiatory act." Those who dominate the discussion fulfill the function of breaking the ice, and of course become highly visible. In short-term groups with no definite structure, what is meant by leadership is not at all clear. Certainly, one prerequisite to being chosen as a leader is standing out and being visible in the group, and which behavior stands out in a group depends on the nature of the situation.

In the observed situation, where, as reported elsewhere (Wilson, 1969), group members uniformly initiated high rates of task-oriented interaction, standing out would be dependent upon sheer volume of interaction. Thus, in the observed condition, high participators were chosen as leaders. In the nonobserved condition, where the groups initiated less task-oriented interaction (see Wilson, 1969), those *S*s whose interaction profiles showed high task orientation would be most visible. As mentioned above, the choice of leaders in the nonobserved groups tended to depend more upon task orientation. In addition, it was found in a separate analysis, not reported here, that the rankings of *S*s on a dimension called "quality of ideas" were significantly more highly related to leadership rankings in the nonobserved condition than in the observed condition. Thus, contributions to the task at hand are necessary to gain leadership in groups not bothered by the problems raised by the pressure of an *O*.

A final possible explanation for the intercorrelations observed is suggested by Gibb (1950) in a study of ad hoc groups, where both *O* and peer ratings of leadership were obtained. He suggests that "participants may have become increasingly aware of the experimental interest in leadership in the study and may have endeavored to make their own records 'look better' by choosing only those whom they expected the observers would also choose on quite different criteria [p. 669]." In the present study, the observed *S*s' ability to verbalize what they felt *O* was looking for indicates that they were searching for what Orne (1962) has called the "demand characteristics" of the situation. Thus, their responses on the questionnaire may have been consciously or unconsciously influenced by what they felt were *O*'s criteria for evaluation. If this were the case, those being most visible and having made ingratiating responses would be seen as ready candidates for leaders.

If only the overtly observed groups had been the source of data, the statistically significant relationships among leadership, participation, and dominance-ascendance would be consistent with the results of earlier studies. However, the failure of these relationships to reach statistical significance in the nonobserved condition plus the fact that in the earlier studies these interrelationships were not generally strong, suggests that these relationships may hold only for groups operating in the presence of an *O*. This strongly suggests that *O* is a situational variable whose effect must be considered when attempting to make generalizations about leadership and group behavior based upon observations made in test or experimental situations.

It is now generally accepted that a leader in one situation will not necessarily be a leader in another. Gibb (1954) suggests that the person who best fulfills the needs of the group members at a particular time will be seen by the others as a leader. This is certainly clear from the present study, where it was shown that when an investigator constructs a situation for the purpose of observing behavior in groups, he is manipulating the context that determines the behavior of various group members, their perceptions of each other, and the criteria used for evaluating leadership.

These results should be of special interest to personnel managers who often attempt to predict leadership behavior in other contexts by observing behavior in a group situation where *O* is a central component. This may lead to erroneous predictions. For example, the person who, through peer nominations, emerges as a leader in an experimental or observational setting is likely to possess qualities that are useless when the individual is operating in situations where strong self-oriented needs cannot be gratified. It is interesting that the leaders in the non-observed conditions appeared to be more task oriented, which seems more in line with the common sense notion of leadership. This type of person was, most likely, performing adequately in the observed condition but merely did not come to light. He would therefore be the type of person who could function well in a wide range of group situations.

In group-leadership studies, where the nature of the situation is of obvious importance, it is impossible to ignore the impact of the research process or the behavior being observed. It is hoped that more studies dealing with the nature of this impact will be forthcoming in the future and that pure and applied researchers will utilize this information to help eliminate errors in making generalizations from experimental test situations.

REFERENCES

BALES, R. F. *Interaction process analysis.* Cambridge, Mass.: Addison-Wesley, 1950.

BASS, B. M. An analysis of the leaderless group discussion. *Journal of Personnel Psychology,* 1949, **33**, 527–537.

BASS, B. M. The leaderless group discussion. *Psychological Bulletin,* 1954, **51**, 485–492.

BASS, B. M. *The Orientation Inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1962.

BASS, B. M. Social behavior and the Orientation Inventory: A review. (Tech. Rep. No. 9) Pittsburgh: University of Pittsburgh, 1965. (Contract No. 624 (14).

BORG, W. R. Predictions of small group role behavior from personality variables. *Journal of Abnormal and Social Psychology,* 1960, **60**, 112–116.

BORGATTA, E. F. Analysis of social interaction and sociometric perception. *Sociometry,* 1954, **17**, 7–32.

CARTER, L. W., HAYTHORNE, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *Journal of Abnormal and Social Psychology,* 1950, **46**, 589–595.

CATTEL, R. B., SAUNDERS, D. R., & STICE, G. F. *The Sixteen Personality Factor Questionnaire.* Champaign, Ill.: The Institute for Personality and Ability Testing, 1951.

CATTEL, R. B., & STICE, G. F. Four formulae for selecting leaders on the basis of personality. *Human Relations,* 1954, **1**, 493–507.

GIBB, C. A. The sociometry of leadership in temporary groups. *Sociometry,* 1950, **13**, 226–243.

GIBB, C. A. Leadership. In G. Lindzey (Ed.), *Handbook of social psychology.* Vol. 2. Reading, Mass.: Addison-Wesley, 1954.

KIRSCHT, J. P., LODAHL, T. M., & HAIRE, M. Some factors in the selection of leaders by members of small groups. *Journal of Abnormal and Social Psychology,* 1959, **58**, 406–408.

MANN, R. D. A review of the relationship between personality and performance in small groups. *Psychological Bulletin,* 1959, **56**, 241–270.

ORNE, T. M. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17**, 776–783.

REDL, F. Group emotion and leadership. In A. P. Hare, E. F. Borgatta, & R. F. Bales (Eds.), *Small groups.* New York: Alfred A. Knopf, 1955.

RIECKEN, H. W. A program for research on experiments in social psychology. In N. F. Washburne (Ed.), *Decisions, values and groups.* New York: Pergamon Press, 1958.

ROSENBERG, M. J. When dissonance fails: Eliminating evaluation anxiety from attitude measurement. *Journal of Personality and Social Psychology,* 1965, **1**, 28–42.

STEPHEN, F. E., & MISHLER, E. G. The distribution of participation in small groups: An exponential approximation. *American Sociological Review,* 1952, **17**, 598–608.

WILSON, S. R. A study of the effects of the observer on small experimental groups under conditions of high and low task saliency. Unpublished doctoral dissertation, Duke University, 1967.

WILSON, S. R. The effects of the laboratory situation on experimental discussion groups. *Sociometry,* 1969, **32**, 220–236.

ZELANY, L. D. Characteristics of group leaders. *Sociology and Social Research,* 1939, **24**, 140–149.

(Received July 7, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 439–442

RELATION OF A TRAINABILITY INDEX TO T-GROUP OUTCOMES[1]

CHRISTOPHER J. PINO[2]

D'Youville College

The ability to screen potential T-group members by a relatively incomplex test procedure could prove to be an asset in the area of human relations training. A trainability index is provided by a combination of scores on Thelen's Reaction to Group Situations Test. Trainability indexes were calculated, and self-disclosures and empathy ratings were measured previous to and following sensitivity training of two groups—psychiatric aides and Office of Economic Opportunity aides. Positive relationships between trainability and the two T-group outcome measures (openness and empathy) were found for both T groups. This provides further validation for the trainability index in human service professions.

The sensitivity training (or T-group) movement has provided the social sciences with the method for instituting personal and social change, as well as a productive channel for small group experimentation. Bradford, Benne, and Gibb (1964) list the following major objectives of the T group: (*a*) to learn more about one's self, (*b*) to facilitate group-member function in other group situations, (*c*) to be a more effective group member, and (*d*) to create greater understanding of group behavior.

In the present study, the T group was seen as a vehicle for developing interpersonal skills necessary to assist mental health aides and Office of Economic Opportunity (OEO) aides in a vital aspect of their job function—that is, in their interpersonal interactions.

The specific purpose of the study was to evaluate an index for predicting individual differences in trainability among participants in such T-group activities in these human service professions. Recent reviews of the effectiveness of T groups have been provided by Buchanan (1967), House (1967), and Campbell and Dunnette (1968). Although the evidence is not uniformly positive, it is clear that positive results can be achieved. There are still a number of questions regarding the conditions under which effective results are obtained. An emerging finding is that different individuals may react to such experiences in different ways.

Steele (1968), for example, attacks the idea that all individuals have adequate skills for learning in T groups. Using the Sensation-Intuition (S-N) scale of the Myers-Briggs type indicator, he found that trainees scoring high on the intuitive (N) mode of perception (i.e., where the perceiver adds to whatever is given in the stimulus situation itself through his ideas and associations) were more effective learners in the T group than trainees scoring at the sensation (S) end of the continuum (i.e., regarding the process of becoming aware of things directly through the S senses). Smith (1966) points out that individual differences are great in the benefits derived from T-group experience.

Portal-Foster (1970) applied Bion's (1951) work-emotionality theory to small discussion groups of school teachers. In nearly all instances, his analysis supported the theory as being a practical descriptive technique. The theory failed, however, in accounting for individual differences under group pressures.

Stock (1964) summarized a number of independent studies regarding individual differences among T-group participants, which showed that certain trainee characteristics were associated with readiness to learn and positive change while other characteristics were associated with lack of positive change.

---

[1] The author would like to thank Lawrence Whisler for his help in collecting data at Chicago State Hospital.

[2] Requests for reprints should be sent to Christopher J. Pino, Department of Psychology, D'Youville College, 320 Porter Avenue, Buffalo, New York 14201.

A major step toward identifying these characteristics is the work of Stock and Thelen (1958). They developed a trainability index derived from a test they devised to tap interpersonal tendencies in T groups—that is, the Reaction to Group Situations Test (RGST). The theoretical underpinnings of this test were derived from Bion's (1951) work-emotionality theory of group behavior. Mathis (1955) conducted a validation study that demonstrated that the trainability index was predictive of self-perceptual changes derived from laboratory training.

Since, in many instances, the T-group approach has been shown to be an effective method for training human relations skills in its participants, the assessment of potential members' trainability would seem a necessary step in the selection process. Stock and Thelen's (1958) RSGT yields a trainability index that may be calculated prior to training to accomplish this task. The present research related this trainability index to two behavioral criteria: openness (cumulative self-disclosure in T groups), and empathy (MMPI Role Play scale). The utility of the trainability index was tested in two different populations: Office of Economic Opportunity (OEO) aides and mental health aides.

METHOD

Subjects

The *Ss* in one of the T groups were 12 Negro, predominantly lower socioeconomic status, mental health (or psychiatric) aide trainees employed at Chicago State Hospital. Most *Ss* had nearly completed high school or its equivalent. All *Ss* had been employed at the hospital for at least 3 mo., none for longer than 1 yr. All were residents of the inner city. Their ages ranged from 18 to 45 yr.

The second T group was composed of 9 OEO aides employed in a rural county in western New York state. All the six Caucasian and three Negro members were high school graduates and were also from the lower and lower-middle socioeconomic level.

Instruments

Self-reference content analysis. Reisel (1959) adopted the self-reference content analysis scheme from Bugental (1948). Self-references during interaction are seen as a method of disclosing self to others. The analysis is a simplified scoring system that can be easily learned. According to Culbert (1968), two phases of the reference system are:

> First there is the determination of thought units. A thought unit is defined as a series of words expressing a single idea or thought; basically, it is equivalent to a simple grammatical sentence. Second is the assignment of thought units to content categories. There are three types of categories having to do with self-references which, for these purposes, imply self-disclosure. The first type of self-reference refers to a person talking exclusively about himself. In this case he is either subject or the subject and object of the thought units, which are assigned to the content category "S→". The second type of self-reference refers to the person discussing the way some group member or incident affects him. Such units are collected under the symbol "←S". The third type of self-reference refers to the person discussing the effect of his participation on the group or on some sub-set of group members. These thought units are collected under the symbol "S→" [pp. 53–54].

All nonself reference statements are given the symbol N. For use in the present research a simplified method was used. Speech segments were assigned either as S (self-reference) or as N (nonself-reference).

Role-Playing scale. The Role-Playing scale of the MMPI (Dahlstrom & Welsh, 1960) is a 31-item true-false scale tapping the ability to exchange roles, or empathic ability. Specifically, McClelland's (1951) definition is that it "refers to the capacity with which a person can perceive and act out organized behaviors and roles" (i.e., putting himself in someone else's position). Split-half reliability of the criterion (multiple student ratings of role-playing ability) was .66; the test-retest reliability was .69 over a 1-mo. period on a second sample. Two check samples were used as further tests of validity.

Reaction to group situation test. The RSGT, developed by Stock and Thelen (1958), translated Bion's (1951) work-emotionality theoretical framework into an objective test. A variety (28) of group situations make up the test items (incompleted sentence stems that *S* fills in). The test is untimed. Stock and Thelen state:

> The subject's responses are scored along three dimensions: acceptance or non-acceptance of the stimulus, the kind of affect introduced overtly or covertly into the response, and the manner of the response. Early work established the reliability of the scoring procedure and demonstrated that descriptions of the same individuals from the R.S.G.T. protocols corresponded with descriptions of the same individuals from observations of their behavior in the group [p. 226].

Mathis (1958) conducted validation studies on a trainability index derived from the RSGT. The trainability index was given as the ratio of adient (approach behaviors) to abient (avoidance behaviors) indicators from the RSGT.

Previous studies with this test (R.S.G.T.) suggested that relative valences for fight, pairing, dependency, and flight (categories derived from the R.S.G.T.) might be obtained as the number of times each of these valences respectively was introduced in the responses of the 28 emotionally charged stimulus items. An estimate of the conflict over each emotionality could be obtained from the balance of accepted to rejected stimuli keyed to each modality (e.g., fight, dependency, etc.); and immobilization could be estimated from the number of times responses in the test gave signs of emotional blocking. Thus, an Index of Trainability would be the sum of fight, pairing, and conflict responses divided by the sum of flight, dependency, and immobilization responses [p. 151].

Procedure

Both groups (OEO aides and mental health aides) were involved in similar human relations training experiences. The T groups met for a 2½–3-hr. session, one morning per week for a 4-mo. period. The same trainer met with both groups. The format was a traditional T group one, heavily group-process oriented, with some structured experiences (e.g., Jones & Pfeiffer, 1969, and Schutz, 1969) interspersed. Some self-rating feedback devices (e.g., Miles, 1959; Pino, 1969), and occasional lecturettes (e.g., "JoHari's Window" of National Training Laboratories, 1965) were also employed.

The RSGT was administered to both groups prior to training. The RSGT protocols were scored by two trained raters (interrater reliability = .83), and trainability indexes were calculated for each member. Each session, for both T groups, was taped, and 10-min. segments from every session were chosen randomly and rated for self-disclosure according to Reisel's (1959) self-reference content analysis system by two trained raters (interrater reliability = .88). Both groups were given the MMPI Role Play scale (a measure of empathy) before and after training.

Hypotheses

1. In each T group, trainability and frequency of cumulative self-disclosure statements will be associated (indicating openness behavior).
2. In each T group, trainability and pre-post-positive changes in MMPI Role Play scale scores will be associated (indicating increases in empathic ability over time).

Statistical Treatment

For each of the T groups, all members were ranked in terms of their trainability index, their pre-post-positive change in Role-Play scale scores, and in regard to cumulative frequency of self-disclosure behavior for the T group sessions over the 4-mo. period. Spearman rank correlation coefficients (Siegel, 1961) were calculated for each group between trainability and cumulative self-disclosure rankings. To test the two hypotheses, the four Spearman correlation coefficients (r_s) were tested for significance by means of the procedure provided by Siegel (1961, p. 211).

RESULTS

The results showed that for both T groups: (*a*) trainability and pre-post change in the Role Play scale are significantly associated (.86 for mental health aides and .67 for OEO community aides), and (*b*) trainability and members' self-disclosure are significantly associated (.91 for mental health aides and .88 for OEO aides). All these coefficients are significant (one-tailed tests) at the .01 level except for the .67 coefficient, which is significant at the .05 level.

DISCUSSION

Sensitivity training is seen as a valuable method of training those interpersonal skills necessary in any helping process. Thus, the job effectiveness of the two groups studied here (community aides and mental health aides) would seem to be predicated on these types of skills. A search of the research literature on T groups indicates that the effectiveness of T groups in building these skills differs according to members' potential trainability.

The research conducted here tested this in terms of two outcome criteria: (*a*) increase in empathy (measured by MMPI role scale), and (*b*) self-disclosure behavior of members cumulated over the training period. Results indicated a strong association between trainability and the outcome measures, providing further validation of the trainability index and demonstrating its utility in two meaningful contexts, that is, the training of mental health aides and OEO community aides.

Mathis (1955) suggested that the trainability index could be a useful screening device to pick trainees for laboratory training. The present research may provide a basis for further studies in assessing the trainability index as a method for selecting candidates for jobs in human services. Since many of the human services job training programs do use laboratory training methods, it follows that the trainability index could play a decisive role in selecting candidates that could profit from the training experiences and, perhaps, sustain changes in helping skills (e.g., empathy, openness, etc.).

There is a backlog of research to indicate that such variables are essential in helping relationships provided by human service workers (Carkhuff, 1970). Mental health aides, in many inpatient settings are traditionally responsible for a majority of staff-patient interactions. There is an increasing use of these nonprofessional workers in both inpatient and outpatient counseling capacities. The OEO aides have a more unstructured role in terms of their helping relationships. The OEO aides are employed in one of the several OEO Community Action Program's components—that is, Head Start, Family Planning, Community Organization, Legal Aide, Neighborhood Youth Corps. Job assignments may range anywhere from a teacher aide to nonprofessional youth counselor, to community organizer, etc. A common thread to all assignments is the requisite of supplying a helping relationship to the economically deprived. Concretely, the job involves supplying information and technical assistance, based on the ability to enter into trust-building human relations, where the consumer (low-income family) views the aide as a competent helper and worthy confidant.

REFERENCES

Bion, W. R. *Experience in groups.* New York: Basic Books, 1951.

Bradford, L., Benne, K., & Gibb, J. *T-group and laboratory method.* Washington, D. C.: National Training Laboratory, 1964.

Buchanan, P. C. Evaluating the effectiveness of laboratory T-group training in industry. In E. A. Fleishman (Ed.), *Studies in personnel and industrial psychology.* Homewood, Ill.: Dorsey Press, 1967.

Bugental, J. F. An investigation of the relationship of conceptual matrix to the self-concept. Unpublished doctoral dissertation, Ohio State University, 1948.

Carkhuff, R. *Helping and human relations.* Vol. 1 & 2. New York: Holt, Rinehart & Winston, 1970.

Campbell, J. P., & Dunnette, M. D. Effectiveness of *T*-group experiences in managerial training and development. *Psychological Bulletin,* 1968, **70**, 73–104.

Culbert, S. Trainer self-disclosure and member growth. *Journal of Applied Behavioral Science,* 1968, **4**, 51–58.

Dahlstrom, W., & Welsh, G. *MMPI handbook.* Minneapolis: University of Minnesota Press, 1960.

House, R. J. *T*-group education and leadership effectiveness: A review of the empiric literature and a critical evaluation. *Personnel Psychology,* 1967, **20**, 1–32.

Jones, J., & Pfeiffer, J. *A handbook of structured experiences for human relations training.* Iowa City: University Associates Press, 1969.

Mathis, A. Development and validation of a trainability index for laboratory training groups. Unpublished doctoral dissertation, University of Chicago, 1955.

Mathis, A. Trainability. In D. Stock & H. Thelen (Eds.), *Emotional dynamics and group culture.* Washington, D. C.: National Training Laboratory, 1958.

McClelland, D. C. A preliminary test of role playing ability. *Journal of Consulitng Psychology,* 1951, **15**, 102–108.

Miles, M. *Learning to work in groups.* New York: Columbia Press, 1959.

NTL Reading Book. Washington, D. C.: National Training Laboratories of the National Education Association, 1965.

Pino, C. J. Rating scale for interpersonal effectiveness. Unpublished manuscript, Clarence, N. Y., 1969.

Portal-Foster, E. An application of work-emotionality theory to adult education. *Adult Education Journal,* 1970, **20**, 3–16.

Reisel, J. A search for behavior patterns. Unpublished doctoral dissertation, University of California at Los Angeles, 1959.

Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1961.

Schutz, W. *Joy.* New York: Evergreen Press, 1969.

Smith, H. C. Sensitivity to people. New York: McGraw-Hill, 1966.

Steele, F. Personality and laboratory style. *Journal of Applied Behavioral Science,* 1968, **5**, 25–45.

Stock, D. Research in sensitivity training. In L. Bradford, K. Benne, & J. Gibb (Eds.), *T group theory and laboratory method.* Washington, D. C.: National Training Laboratory, 1964.

Stock, D., & Thelen, H. *Emotional dynamics and group culture.* Washington, D. C.: National Training Laboratory, 1958.

(Received July 22, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 443–448

SIMILARITY OF POST-INTERVIEW TRAIT RATING INTERCORRELATIONS AS A CONTRIBUTOR TO INTERRATER AGREEMENT IN A STRUCTURED EMPLOYMENT INTERVIEW[1]

MILTON D. HAKEL[2]

Ohio State University

Degree of interview structure, use of scaled-expectation rating scales, and similarity of postinterview trait rating intercorrelations are three variables that presumably influence interrater agreement in employment interviews. Nine nursing interviewers sat as a selection board and interviewed and independently rated 54 senior nursing students. The interviews were highly structured, and ratings were recorded on scaled-expectation scales for general staff nursing positions. Although all interviewers shared essentially the same structure among their postinterview trait ratings, interrater agreement was no better than in previous studies. This finding shows the power of halo in the interview setting. Sharing the same structure in ratings (and using the scaled-expectation scales in a highly structured interview) by no means standardized perceptions of job applicants.

Instrumentation in the selection and placement of job applicants was revolutionized after World War I by the widespread introduction of tests of various abilities and, then, tests of interests and character. These instruments have achieved an important and valuable place in the employment process and are the objects of continuing refinement and calibration. In the last few years, it has been realized that the employment interview, also an instrument for making observations, ought to be given the same attention devoted to employment testing. The focus now is on studying the interviewing process, making it too the object of continuing refinement and calibration.

It takes little acquaintance with the literature on the employment interview to realize that interrater agreement is too low to justify psychometric use of the interview (see reviews by Mayfield, 1964; Ulrich & Trumbo, 1965; Wagner, 1949; and Wright, 1969).

The outlook is not entirely bleak, however. The literature offers suggestions about variables controlling interrater agreement in the employment interview. First, it is generally felt that a structured or patterned interview is more reliable than an unstructured one (Bingham, Moore, & Gustad, 1959; Wagner, 1949; Wright, 1969). Although no study has been reported that directly varied the amount of structure to observe its effects on reliability, studies reported by McMurry (1947) seem to support this view.

A second variable that may control interrater agreement is rating scale format. Maas (1965) showed that agreement on scaled-expectation rating scales is significantly superior to agreement on adjective rating scales anchored in the following manner: very good, good, average, poor, and very poor.

Finally, one finding of the McGill studies (Webster, 1964), that interviewers apparently make selection decisions by testing the goodness of fit of each applicant against a stereotype of an ideal applicant, is especially relevant to interrater agreement. Interviewers sharing similar stereotypes presumably should show higher interrater agreement than interviewers whose stereotypes differ. Intuitively, it seems clear that interviewers who select applicants by reference against dissimilar standards should show larger disagreements than interviewers who use the same standard. The usefulness of this suggestion clearly depends on finding differences among interviewers in their stereotypes for particular occupations. The McGill studies seem to show that all interviewers within an occupation use the same stereotype, but other research by Mayfield and Carlson (1966) and Hakel, Hollmann, and Dunnette (1970) questions this position, thus making an analysis of stereotype similarity and interrater agreement desirable.

---

[1] Marvin Dunnette's guidance in this study is gratefully acknowledged.

[2] Requests for reprints should be sent to Milton D. Hakel, Department of Psychology, Ohio State University, 1945 North High Street, Columbus, Ohio 43210.

This research examines interrater agreement as a function of rater similarity of within-interviewer postinterview trait rating intercorrelations, recovered from ratings made on scaled-expectation rating scales in highly structured board interviews. The procedures used, such as a highly structured interview guide, scaled-expectation scales, the board interview setting, and a preexperiment lecture on biases (leniency and halo) in ratings were all intended to facilitate agreement among raters. Thus, the study really focuses on one question: Do interviewers who structure their postinterview ratings similarly agree more with each other than with other interviewers whose structures differ?

It is mathematically necessary that if two raters show perfect agreement on all scales, then the within-rater intercorrelations will be identical for those two raters. But this relation is not symmetrical. Two raters might have identical patterns of trait intercorrelations and show no agreement whatsoever ($r = .00$) or even perfect disagreement ($r = -1.00$). This research is directed at the latter point.

METHOD

Subjects

Nine nursing administrators and supervisors from seven hospitals served as interviewers in this research. Their median age was 44 yr., with a range of 27–53 yr. All were registered nurses, and seven of them held college degrees in Nursing Administration. Median interviewing experience was 4 yr., with a range of 1–12 yr. All occupied positions in which interviewing nursing job applicants was an integral part of the job.

Fifty-four student nurses from three nursing schools volunteered to serve as general staff nurse job applicants and to be questioned by the interviewing board. All students were in the final 3 mo. of training and were preparing for the Registered Nurse examinations. The investigator met with groups of students from each of the schools to recruit volunteers: the reasons for studying interrater agreement and a description of the board interview procedures constituted the content of these meetings. Rate of participation was high. Each student received a copy of a general staff nurse job description and was asked to study it prior to her interview. This was done to assure that students from each school had relatively common expectations about the job behaviors required of general staff nurses. Each student also completed an application form. The student's median age was 20 yr., with a range of 19–31 yr. Eleven students were interviewed the first day, 15 the second, and 14 on both the third and fourth days. Interviews lasted about 15 min. each.

Procedures

The interviews proceeded as follows: copies of the first interviewee's application form were distributed to each interviewer for a brief examination. The applicant entered the room, was introduced to the board, and was seated at the head of the table. Questioning proceeded according to an interview guide, which was developed and revised with the cooperation of the participating interviewers for use in this study. The final version of the guide contained nine questions (one asked by each interviewer).

The nine questions were as follows: (*a*) On your application blank, you described your idea of a good supervisor as being a person who. . . . Why do you feel this way? On your application, you described a poor supervisor as one who. . . . Why do you feel this way? (*b*) Suppose you're in charge of 10 bed patients, 4 of them are acutely ill, and you have only one aide to help you. How would you organize the work to be done? (*c*) Suppose you are on duty and extremely busy. One of your patients calls you and begins to complain again about the care he has received. What would you say to him? What would you do? (*d*) Do you feel that it is important to keep all your charts and records up to date? Why? (*e*) An aide has just completed giving morning care to a patient and you come into his room to check his condition. What kinds of things would you look for? (If no satisfactory answer is made, ask, "What would you look for in terms of the patient's comfort?") (*f*) What would you do if a patient who had just received an injection of penicillin told you he was feeling queer and having trouble breathing? (*g*) Suppose you are caring for a patient who is in critical condition. The patient's family is at the hospital, they are quite upset and know very little about the patient's current condition. What, if any, is your responsibility to the family? If the patient's wife stops you in the hall to ask about her husband's condition, what would you say? (*h*) Nurses have many responsibilities to their patients. One of these responsibilities is simply to possess a sufficient level of technical competence to be able to meet the medical needs of patients. Another of these responsibilities is to render some kind of psychological support to the patient. With respect to your nursing education, in which of these areas of responsibility do you feel you are best prepared? Are you satisfied with your preparation in these areas? (*i*) One last question: In terms of your nursing career, what would you like to be doing 10 yr. from now?

Probing and follow-up questions were encouraged. The emphasis throughout the interview was on obtaining as much information about the applicant as possible, and to this end the guide contained at least one question specifically relevant to each of the scaled-expectation rating scales used in the study. Following completion of the questioning, the applicant left the room, and the interviewers made independent ratings. The second application form was distributed, the applicant was introduced, and so on. After about half of the interviews on each day, questions on the interview guide were systematically reassigned in an attempt to equate interviewer use of the questions over the entire series of interviews.

At the end of an interview, each interviewer independently rated the applicant on six rating scales. The scales were titled Conscientiousness, Observational Ability, Skill in Human Relationships, Organizational Ability, Knowledge and Judgment, and Expected Level of Performance. They were presented in this order, one scale to a page, in a rating booklet. The first five of these scales were developed by Smith and Kendall (1963) and the National League for Nursing for the evaluation of general staff nursing performance and are known as scaled-expectation scales. The noteworthy feature of these scales is the use of behavior examples to define levels of performance on the scale. The use of behavior examples to anchor scales virtually eliminates the semantic ambiguity that attends anchors such as "average," "superior," and similar adjectival anchors. The elimination of such semantic ambiguity in the scale anchors is reflected in the scale reliabilities, the lowest of which is .97 (Smith & Kendall, 1963). The availability of these behaviorally defined scales dictated that nursing interviewers and student nurses be *S*s for this investigation. The sixth scale, Expected Level of Performance, was developed especially for this study, and was used to secure a rating of overall suitability. While its format was similar to the scaled-expectation format, the sixth scale differed in that summary predictions (e.g., I am certain that this applicant will be an outstanding nurse) rather than behavior examples anchored the scale. Additional details of the procedures are given by Hakel (1966).

RESULTS

The analyses reported here were completed in the framework of a double cross-validation design to assess the stability of each interviewer's ratings. There would be little sense in examining differences between interviews if their ratings were unstable. Accordingly, the sample was split into two cross-validation groups of 27 each.

Each interviewer's ratings on the five scaled-expectation scales and the overall suitability scale were intercorrelated within each cross-validation group.[3] The within-rater intercorrelations are uniformly high: Of the 270 correlations between traits within raters within groups, about 80% of them reached or exceeded the 5% significance level. Except for one interviewer, there was a general failure to discriminate empirically among the traits being rated. Stated differently, all raters (except Number 3) seemed to have structured their ratings similarly.

To examine the stability of each rater's intercorrelations and to test for similarity between raters in the magnitudes and configurations of their intercorrelations, multiple-regression equations relating the scaled-expectation scales to the overall ratings were computed in each cross-validation group for each interviewer (18 equations in all). Once developed, these equations were crossed to their companion samples, and very little shrinkage was observed: The median developmental *R* was .88 (with a range of .79–.92) and the median cross-validated *r* was .83 (with a range of .58–.90). Next, these equations were applied to every other interviewer's sets of ratings in a validity extension study. If Interviewer I's regression equation could predict Interviewer J's ratings of overall suitability as accurately as it could predict I's ratings of overall suitability, then I and J must have structured their ratings similarly. Given large shrinkage, the structures must have been different (since the equations were stable). The validity extension correlations are shown in Table 1.

Every interviewer's ratings of overall suitability can be predicted as well (and sometimes even better) by another's equation as by

---

[3] These intercorrelations, as well as means and standard deviations for each interviewer on each scale, and interrater agreement correlations for the scaled-expectation scales were computed and have been deposited with the National Auxiliary Publications Service. Order Document No. 01559 from the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

TABLE 1

CORRELATIONS OBTAINED WHEN EVERY REGRESSION EQUATION IS APPLIED TO EVERY SET OF RATINGS TO PREDICT RATINGS ON EXPECTED LEVEL OF PERFORMANCE

Ratings predicted	By equation																	
	1A	1B	2A	2B	3A	3B	4A	4B	5A	5B	6A	6B	7A	7B	8A	8B	9A	9B
1A	*89*	*88*	82	88	87	87	88	87	81	75	86	83	84	87	87	80	85	86
1B	*88*	*89*	85	88	84	87	87	86	84	71	83	81	85	88	87	76	85	85
2A	84	85	*89*	*85*	86	88	87	87	80	82	87	86	88	86	85	84	87	86
2B	87	86	*82*	*87*	85	85	86	84	83	70	84	84	83	86	86	80	85	86
3A	63	60	67	66	*80*	*70*	67	57	44	42	76	66	65	67	56	64	71	69
3B	62	60	57	63	*62*	*70*	69	65	44	38	57	59	66	65	52	39	58	57
4A	82	80	76	82	77	83	*83*	*83*	72	73	73	78	82	83	80	56	73	81
4B	87	88	84	86	79	89	*90*	*91*	81	82	77	83	88	89	84	64	78	82
5A	69	69	65	67	65	67	69	69	*72*	*58*	64	64	67	69	66	59	65	64
5B	72	71	74	73	73	75	75	75	*64*	*80*	72	74	75	72	73	66	70	75
6A	86	86	88	88	91	87	87	83	70	63	*92*	*83*	84	86	86	86	90	86
6B	77	74	72	79	79	79	79	76	59	66	*76*	*82*	76	76	77	67	72	89
7A	74	71	72	75	80	85	83	84	56	46	81	79	*88*	*81*	65	54	77	74
7B	88	87	81	86	87	87	88	87	76	54	85	85	*87*	*90*	78	69	86	81
8A	88	88	84	89	87	87	87	86	81	79	86	85	84	86	*89*	*84*	86	88
8B	83	81	84	84	82	84	83	80	67	75	87	83	81	82	*84*	*89*	86	86
9A	79	78	79	78	79	79	79	77	73	67	80	78	78	78	77	75	*80*	*77*
9B	80	79	79	81	81	81	81	79	71	74	81	89	80	89	81	78	*80*	*82*

Note.—Entries in the "ratings predicted" column identify the data set to which the regression equations were applied. Thus, 1A in the "ratings predicted" column indicates that Interviewer 1's ratings of applicants in Group A were the data on which the correlations reported in that row were obtained. The correlation in the upper left-most cell of the table, .89, was obtained by applying Equation 1A (developed on ratings of Group A by Interviewer 1) to data set 1A. Foldback and cross-validated correlations are italicized. All other entries are validity extension correlations. Decimals have been omitted.

his own. The range of the validity extension correlations is .38–.91 with a median of .82 (recall that the median cross-validated r is .83). Moreover, all except five of the 288 validity extension correlations reach or exceed the 1% significance level (all exceed the 5% level). Nearly 70% of these correlations reach or exceed a value of .75. Considering the high magnitudes of the developmental, cross-validation, and validity extension correlations, and recalling the hypothesized relationship between interrater agreement and structural similarity of postinterview ratings, it would be expected that interrater agreement should be quite high.

Between-interviewer correlations were computed on the overall suitability scale,[4] and are presented in Table 2. The median interrater agreement correlation is .34, with a range of −.20–.72. While about 40% of these correlations reach or exceed the 5% significance level, it is obvious that interrater agreement is low. The agreement correlations reported in Table 2 are no higher, but no lower, than the correlations reported by Wagner (1949) and Ulrich and Trumbo (1965), despite these controls and precautions: All interviewers had access to the same information from each applicant, ratings were made on extremely well-constructed, unambiguous rating scales, and all interviewers shared essentially the same pattern of postinterview ratings on the six scales.

DISCUSSION

At the outset of this investigation it was known that the interrater agreement correlations would be mathematically independent of the similarity of intercorrelations among postinterview ratings. It was expected, however, that they would not be quite so independent empirically. The outcome of "run-of-the-mill" agreement correlations despite structured interviews, unambiguous rating scales, and structural similarity of ratings was not at all expected. Thus, additional analyses were made to rule out several potential explanations of these findings.[5] These analyses showed that the observed results (high structural similarity and low interrater agreement) are attributable wholly to halo in the ratings.

---

[4] The median agreement correlations for each SE scale are as follows: Conscientiousness, .21; Observational Ability, .31; Skill in Human Relations, .16; Organizational Ability, .33; and Knowledge and Judgment, .36.

What can be said about interrater agreement? Clearly, the procedures used in this study did not foster higher agreement than has been found in other studies (Wagner, 1949). Structural similarity of the postinterview ratings on the six scales was exhaustively studied (a stronger test could be provided by using more rating scales), but the brute fact is that sharing the same structure in ratings by no means standardizes perceptions of job applicants. This latter conclusion fits particularly well with Passini and Norman's (1966) discovery of a highly stable 5-factor structure underlying peer nominations made by total strangers. Thus, in studying the comparability of judges (raters, interviewers, clinicians, etc.) it is not sufficient to show that they have similar statistical functions (e.g., regression equations) operating in their judgments, but that they agree in their judgments on a given set of stimuli.

The value of the scaled-expectation scales for enhancing interrater agreement was far less than expected, especially given Maas's (1965) report of median agreement correlations of .55 and .69 in two studies.

TABLE 2

INTERRATER AGREEMENT CORRELATIONS FOR RATINGS OF EXPECTED LEVEL OF PERFORMANCE

Interviewer	1	2	3	4	5	6	7	8	9
1		41	22	58	−20	43	43	26	32
2	05		28	27	14	59	41	23	47
3	32	07		31	−04	45	38	01	12
4	51	11	19		−03	36	34	50	11
5	63	14	52	61		02	15	37	32
6	57	03	25	59	63		34	35	37
7	30	25	60	40	58	23		30	47
8	50	16	28	51	64	72	23		18
9	36	25	38	29	63	18	43	41	

Note.—Nine interviewers, 27 applicants in each group. Group A correlations are below diagonal; Group B are above diagonal. Decimals have been omitted.

Did these interviewers share the same stereotype of an effective general staff nurse? They certainly shared similar patterns of postinterview trait rating intercorrelations among the six rating scales, but it does not follow necessarily that they shared the same stereotype. This leaves the possibility that interrater agreements may be related to similarity of stereotypes provided that different procedures (such as those used by, e.g., Hakel, Hollmann, & Dunnette, 1970) are adopted for measuring stereotypy.

In the last analysis these data seem to reaffirm the importance of the first major finding of the McGill (Webster, 1964) investigations. Springbett found that first impressions are remarkably potent, and that in the vast majority of interviews (73% overall), the first impression is also the final one (Webster, 1964). If first impressions are so stable, then the results of this investigation could be the consequence of interrater disagreement on first impressions. Thus, it would be desirable to make an analysis of both agreement on first impressions and also changes in agreement over the course of the interview.

REFERENCES

BINGHAM, W. V. D., MOORE, B. V., & GUSTAD, J. W. *How to interview.* (4th ed.) New York: Harper & Sons, 1959.

HAKEL, M. D. Perceiver differences in interpersonal perceptions: An analysis of interrater agreement on scaled-expectation rating scales in an employment interview setting. Unpublished doctoral dissertation, University of Minnesota, 1966.

HAKEL, M. D., HOLLMANN, T. D., & DUNNETTE, M. D. Accuracy of interviewers, Certified Public Accountants, and students in identifying the interests of accountants. *Journal of Applied Psychology,* 1970, **54**, 115–120.

MAAS, J. B. Patterned scaled-expectation interview: Reliability studies on a new technique. *Journal of Applied Psychology,* 1965, **49**, 431–433.

MAYFIELD, E. C. The selection interview—A re-evaluation of published research. *Personnel Psychology,* 1964, **17**, 239–260.

---

[5] Details on these post hoc analyses, done to examine (*a*) capitalizing on chance in small-sample multiple regression, (*b*) systematic (odd-even) versus random assignment to cross-validation groups, (*c*) application of the Spearman-Brown prophesy formula to a rating situation, (*d*) comparability of factor analytic and multiple-regression findings, and (*e*) the influence of halo error in ratings (taken as the first unrotated principal components factor) on agreement, are given in Hakel (1966).

MAYFIELD, E. C., & CARLSON, R. Selection interview decisions: First results from a long-term research project. *Personnel Psychology,* 1966, **19**, 41–53.

MCMURRY, R. N. Validating the patterned interview. *Personnel,* 1947, **23**, 263–272.

PASSINI, F. T., & NORMAN, W. T. A universal conception of personality structure? *Journal of Personality and Social Psychology,* 1966, **4**, 44–49.

SMITH, P. C., & KENDALL, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology,* 1963, **47**, 149–155.

ULRICH, L., & TRUMBO, D. The selection interview since 1949. *Psychological Bulletin,* 1965, **63**, 100–116.

WAGNER, R. The employment interview: A critical summary. *Personnel Psychology,* 1949, **2**, 17–46.

WEBSTER, E. C. *Decision making in the employment interview.* Montreal: McGill University Industrial Relations Centre, 1964.

WRIGHT, O. R., JR. Summary of research on the selection interview since 1964. *Personnel Psychology,* 1969, **22**, 391–413.

(Received July 1, 1970)

Manuscripts Accepted for Publication in the *Journal of Applied Pyschology*

Do Cultural Differences Affect Job Satisfaction?: John W. Slocum, Jr.*: College of Business Administration, Pennsylvania State University, 223 Boucke Building, University Park, Pennsylvania 16802.

Employment Interviewers' Errors in Processing Positive and Negative Information: Thomas D. Hollmann,* Department of Psychology, Wayne State University, Detroit, Michigan 48202.

Relationship between Intrinsic Needs and Women's Persistence at Work: Elchanan I. Meir*: Department of Psychiatry, Tel-Aviv University, Tel-Aviv, Israel.

Effects of Instructions on Measures of State and Trait Anxiety in Flight Students: Steven F. Bucky,* Charles D. Spielberger, and Ronald M. Bale: Head, Clinical Psychology, Naval Aerospace Medical Institute, Naval Aerospace Medical Center, Pensacola, Florida 32512.

Value of Peer Nominations in Predicting Life Insurance Sales Performance: Eugene C. Mayfield*: Director of Manpower Research, Life Insurance Agency Management Association, 170 Sigourney Street, Hartford, Connecticut 06105.

Hierarchies of Scientist Goal Objects: Individual Characteristics and Performance Correlates: Michael W. McCarrey,* and Shirley A. Edwards: Personnel Psychology Service and Research Centre, Public Service Commission of Canada, Tower "A" Place de Ville, Ottawa, Ontario, Canada.

Organizational Climate: Individual Preferences and Organizational Realities: Benjamin Schneider*: Department of Psychology, University of Maryland, College Park, Maryland 20742.

Immediate Aftereffects of the Finnish Sauna upon Psychomotor Performance and Mood: Jorma Kuusinen,* and Markku Heinonen: Department of Psychology, University of Jyvaskyla, 40100 Jyvaskyla 10, Finland.

Use of Subjective Predictors in Regression Analysis for Policy Capturing: Michael E. Doherty,* and Stuart M. Keeley: Associate Professor, Department of Psychology, Bowling Green State University, Bowling Green, Ohio 43403.

A Longitudinal Study of the Effect of a Human Relations Training Program on Managerial Effectiveness: Herbert H. Hand, and John W. Slocum, Jr.*: College of Business Administration, Pennsylvania State University, 223 Boucke Building, University Park, Pennsylvania 16802.

Children's Reactions to Second-Hand Tobacco Smoke: Paul Cameron*: Department of Psychology, University of Louisville, Louisville, Kentucky 40208.

Effects of Perceptions of Equity and Inequity on Worker Performance and Satisfaction: Robert D. Pritchard,* Marvin D. Dunnette, and Dale O. Jorgenson: Department of Psychology, Purdue University, Lafayette, Indiana 47907.

Effects of Alchohol Ingestion on Tracking Performance during Angular Acceleration: William E. Collins,* Richard D. Gilson, David J. Schroeder, and Frederick E. Guedry, Jr.: Chief, Psychology Laboratory, AC-11, Civil Aeromedical Institute, Federal Aviation Administration, P. O. Box 25082, Oklahoma City, Oklahoma 73125.

---

* Asterisk indicates author for whom address is supplied.

Journal of Applied Psychology
1971, Vol. 55, No. 5, 449–457

TAYLOR–RUSSELL TABLES FOR DICHOTOMOUS CRITERION VARIABLES

NORMAN M. ABRAHAMS,[1] EDWARD F. ALF,[2] AND JOHN H. WOLFE

Naval Personnel and Training Research Laboratory, San Diego

The Taylor-Russell tables have been widely used to evaluate the usefulness of selection tests. They are inappropriate, however, for evaluating predictors used to discriminate discrete normally distributed criterion groups where a point-biserial correlation coefficient is the proper index of validity. This article provides a set of tables, similar to the Taylor-Russell tables, for evaluating the usefulness of selection tests where the assumptions require point-biserial validity coefficients. A graphic illustration of the rationale for developing the tables is also presented.

Taylor and Russell (1939) denounced the practice of evaluating the effectiveness of predictor variables solely on the basis of their validity coefficients. In this significant publication they stressed the critical role of base rates of satisfactory criterion performance and selection ratios in conjunction with validities in assessing the practical significance of predictors. To facilitate this type of evaluation, they prepared tables for determining the effectiveness of a selection test by estimating the proportion of selectees that are considered satisfactory when the test validity, base rate, and selection ratio are known or assumed.

The Taylor-Russell tables were developed from normal bivariate correlation surfaces assuming that the "satisfactory-unsatisfactory" categorization represents a dichotomization of an underlying normally distributed criterion variable. Thus, the proper index of the tests' validity would be either a product-moment correlation computed on a continuous criterion distribution or a biserial correlation computed on a dichotomized underlying normally distributed criterion.

There are times, however, when the criterion variable represents either a true dichotomy or two distinct continuous normal distributions. Under these conditions, of course, the assumption of a single underlying normal distribution required for biserial r is not met, and point-biserial r is a more appropriate statistic. For example, in evaluating the effectiveness of an occupational interest measure in discriminating between two distinct occupational groups, such as mechanics and librarians, it would be hazardous to assume the two groups represent an underlying normally distributed continuum. Similarly, in evaluating masculinity-femininity scales, or scales designed to discriminate clinical groups, the assumptions of point-biserial r may be more appropriate than those of biserial r.

In such cases, where a point-biserial r is required, the Taylor-Russell tables are not appropriate for assessing the effectiveness of a predictor. If point-biserial correlations are used to enter the Taylor-Russell tables, the estimated percentage of satisfactory selectees will be in error. In general, the further the selection ratio and base rate are from .50, the less accurate the estimate of percentage satisfactory becomes. For example, if one entered the Taylor-Russell tables assuming a base rate and selection ratio of .50 and a point-biserial r of .60, the estimated percentage of satisfactory selectees is .70. In reality, however, the correct estimate is .77 when the validity is a point-biserial. Assuming a base rate and selection ratio of .10 and a validity of .60, the percentage satisfactory from Taylor-Russell is .39, in contrast to the actual value of .73. The extent to which the Taylor-Russell tables have been used inappropriately with point-biserials is, of course, unknown.

The present article provides a set of tables parallel to the Taylor-Russell tables for use with point-biserial correlations and the rationale for their development.

A graphic illustration of the method of determining entries in the original Taylor-Russell

---

[1] Request for reprints should be sent to Norman M. Abrahams, Personnel Measurement Research Department, Naval Personnel and Training Laboratory, San Diego, California 92152.

[2] Also at San Diego State College.

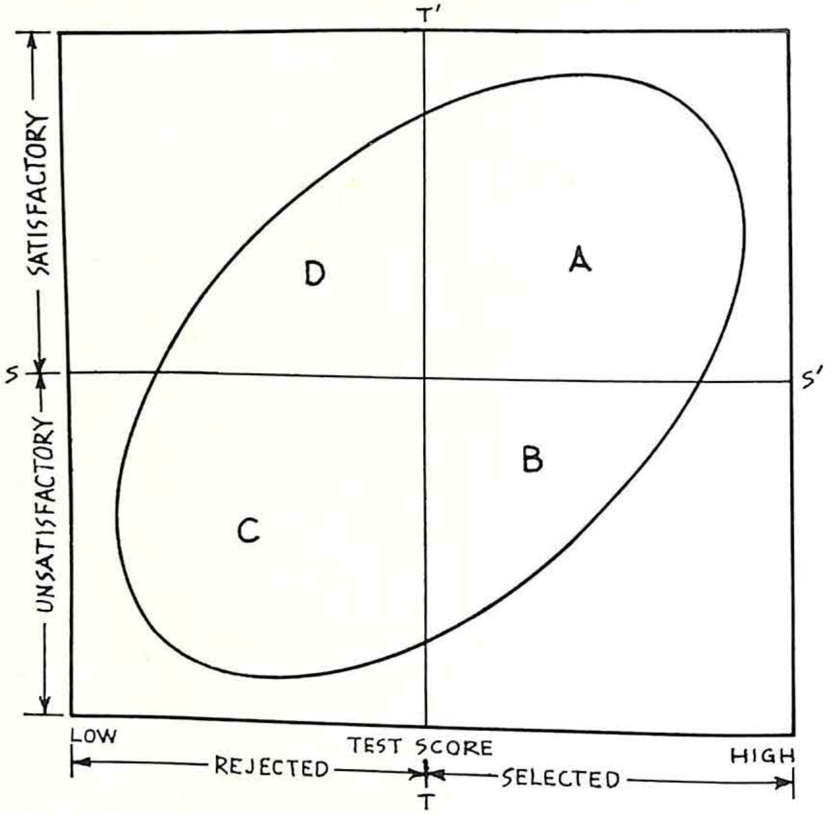


FIG. 1. Theoretical normal bivariate distribution divided into four categories by dichotomizing the test and criterion score distributions.

tables is given in Figure 1. This figure depicts a bivariate plot for a selection test with a validity of .50. The horizontal axis, SS′, represents a particular base rate of satisfactory criterion performance, in this case, 50%. The vertical axis, TT′, denotes the cutting score for a 50% selection ratio. Thus, Quadrant A contains satisfactory employees above the cutting score (true positives), Quadrant B contains unsatisfactory employees above the cutting score (false positives), Quadrant C contains unsatisfactory employees below the cutting score (true negatives), and Quadrant D contains satisfactory employees below the cutting score (false negatives). The entry in the Taylor-Russell tables for this particular combination of validity, base rate, and selection ratio is the proportion $A/(A+B)$, or the "true positive" hit rate.

An analogous illustration for point-biserial is depicted in Figure 2. In this case the validity, represented by a mean difference of 1.34 standard deviation units, is a point-biserial r of .55; the base rate is 50%, as indicated by the equal-sized upper and lower distributions; and the selection ratio, represented by the vertical axis TT′, is 60%. For this particular point-biserial r, base rate, and selection ratio the expected proportion of satisfactory selectees, that is, $A/(A+B)$, is determined and becomes the entry in the new tables.

For each point-biserial r, ranging from .00 to 1.00, and base rate, ranging from .05 to .95, the separation of group means in terms of standard deviation units was determined. In effect, a combined distribution was developed and each selection ratio, varying from .05 to .95, was imposed. For each combination of point-biserial validity, base rate, and selection ratio, the proportion of successful selectees, $A/(A+B)$, was computed. These values are provided in Table 1.

There are 11 segments, corresponding to 11 values (.05, .10, .20, .30, .40, .50, .60, .70, .80, .90, and .95) of proportion of employees considered satisfactory in Table 1. The columns of

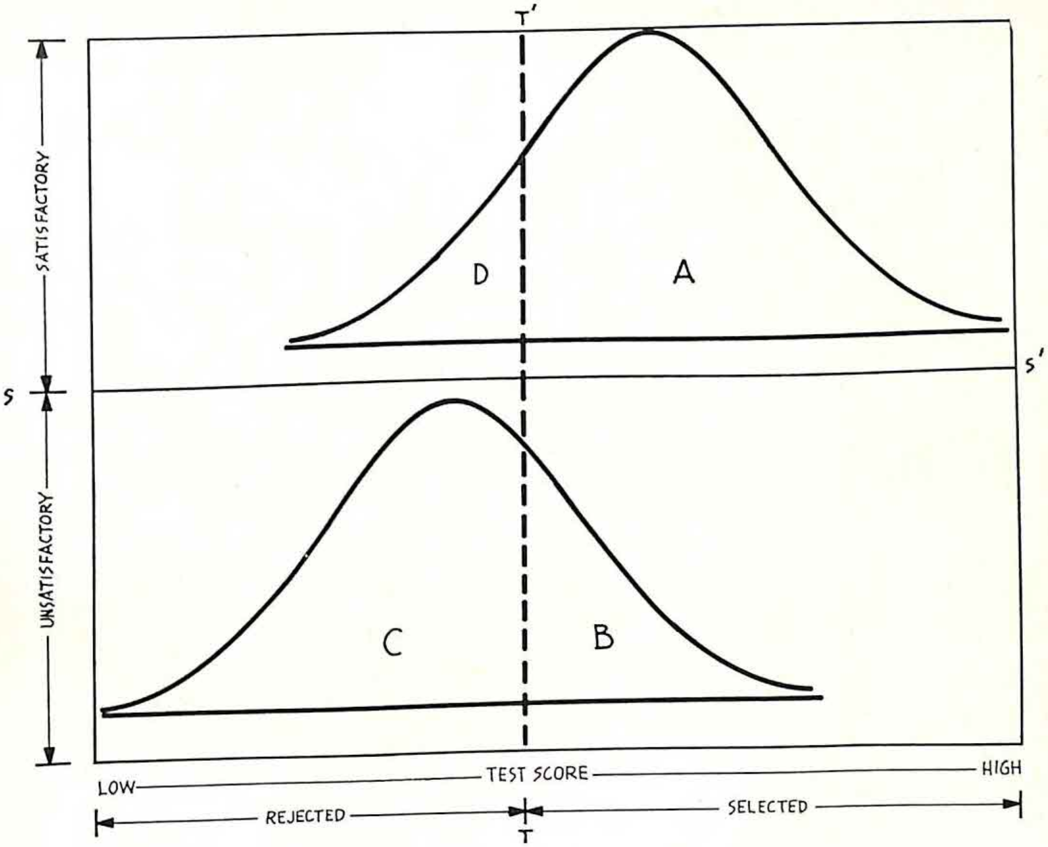


FIG. 2. Theoretical normal univariate distributions of two discrete criterion groups each divided into two categories by dichotomizing the test score distribution.

each segment represent 11 values of the selection ratio: .05, .10, .20, .30, .40, .50, .60, .70, .80, .90, and .95. The rows of each segment correspond to values of point-biserial r from .00 to 1.00 by intervals of .05. The entries in the table indicate the proportion satisfactory among those selected.

To illustrate the use of the table, recall the previously cited example with a point-biserial r of .55, a selection ratio of .60, and a base rate of .50. To obtain the expected proportion of satisfactory selectees, first locate the sixth segment in which the proportion of employees considered satisfactory is .50. Next, locate the row for a point-biserial r of .55 by looking down the first column. Finally, locate the column for a selection ratio of .60, and read the entry at the intersection of this row and column. For this example, the expected proportion of satisfactory selectees is .696.

In some instances it may be valuable to determine the proportion of satisfactory employees below the cutting score of a particular selection ratio, that is, $D/(D + C)$, or various other proportions of interest such as $(A + C)/(A + B + C + D)$.

Once the proportion of satisfactory selectees is determined, the proportion of the total sample falling into each of the four quadrants can be found by the following equations:

$$A = (\mathrm{PS})(\mathrm{SR}),$$
$$B = (\mathrm{SR}) - \mathrm{A},$$
$$C = 1 - (\mathrm{BR} + \mathrm{B}), \text{ and}$$
$$D = (\mathrm{BR}) - \mathrm{A}$$

where:

A–D represent the proportion of cases in each of the four areas shown in Figure 2,

PS = proportion satisfactory among those selected as indicated by the table,
BR = base rate of satisfactory performance, and
SR = selection ratio.

TABLE 1

PROPORTION WHO WILL BE SATISFACTORY AMONG THOSE SELECTED, FOR GIVEN VALUES OF THE PROPORTION OF PRESENT EMPLOYEES CONSIDERED SATISFACTORY, THE SELECTION RATIO, AND POINT-BISERIAL r

r_{pb}	Selection ratio										
	.050	.100	.200	.300	.400	.500	.600	.700	.800	.900	.950
Proportion of employees considered satisfactory = .05											
.00	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050	.050
.05	.076	.072	.066	.063	.061	.059	.057	.055	.054	.052	.051
.10	.112	.099	.086	.078	.072	.067	.063	.060	.056	.053	.052
.15	.157	.131	.106	.092	.082	.075	.068	.063	.058	.054	.052
.20	.213	.169	.128	.107	.092	.082	.073	.066	.060	.055	.052
.25	.278	.210	.151	.121	.101	.087	.076	.068	.061	.055	.052
.30	.352	.255	.173	.133	.109	.092	.079	.070	.062	.055	.053
.35	.433	.302	.194	.144	.115	.095	.081	.070	.062	.056	.053
.40	.519	.348	.212	.153	.119	.097	.082	.071	.062	.056	.053
.45	.607	.393	.227	.159	.122	.099	.083	.071	.062	.056	.053
.50	.694	.431	.238	.163	.124	.100	.083	.071	.062	.056	.053
.55	.776	.462	.245	.165	.124	.100	.083	.071	.062	.056	.053
.60	.849	.484	.248	.166	.125	.100	.083	.071	.062	.056	.053
.65	.909	.495	.250	.167	.125	.100	.083	.071	.062	.056	.053
.70	.954	.499	.250	.167	.125	.100	.083	.071	.062	.056	.053
.75	.982	.500	.250	.167	.125	.100	.083	.071	.062	.056	.053
.80	.996	.500	.250	.167	.125	.100	.083	.071	.062	.056	.053
.85	.996	.500	.250	.167	.125	.100	.083	.071	.062	.056	.053
.90	.996	.500	.250	.167	.125	.100	.083	.071	.062	.056	.053
.95	1.000	.500	.250	.167	.125	.100	.083	.071	.062	.056	.053
1.00	1.000	.500	.250	.167	.125	.100	.083	.071	.062	.056	.053
Proportion of employees considered satisfactory = .10											
.00	.100	.100	.100	.100	.100	.100	.100	.100	.100	.100	.100
.05	.134	.129	.122	.118	.115	.112	.110	.107	.105	.103	.102
.10	.176	.162	.147	.137	.130	.124	.118	.114	.109	.105	.103
.15	.226	.200	.173	.157	.145	.135	.127	.120	.113	.107	.104
.20	.284	.243	.202	.177	.160	.146	.135	.125	.116	.108	.104
.25	.349	.290	.232	.198	.175	.157	.142	.130	.119	.109	.104
.30	.422	.343	.264	.219	.189	.166	.148	.133	.121	.110	.105
.35	.502	.399	.297	.240	.202	.175	.153	.136	.122	.110	.105
.40	.588	.460	.330	.260	.214	.182	.158	.139	.123	.111	.105
.45	.676	.525	.364	.278	.225	.188	.161	.140	.124	.111	.105
.50	.765	.592	.397	.295	.234	.193	.164	.142	.125	.111	.105
.55	.847	.661	.427	.309	.240	.196	.165	.142	.125	.111	.105
.60	.917	.731	.454	.320	.245	.198	.166	.143	.125	.111	.105
.65	.966	.800	.476	.327	.248	.199	.166	.143	.125	.111	.105
.70	.991	.864	.490	.331	.249	.200	.167	.143	.125	.111	.105
.75	.999	.920	.497	.333	.250	.200	.167	.143	.125	.111	.105
.80	1.000	.964	.500	.333	.250	.200	.167	.143	.125	.111	.105
.85	1.000	.990	.500	.333	.250	.200	.167	.143	.125	.111	.105
.90	1.000	.999	.500	.333	.250	.200	.167	.143	.125	.111	.105
.95	1.000	.999	.500	.333	.250	.200	.167	.143	.125	.111	.105
1.00	1.000	1.000	.500	.333	.250	.200	.167	.143	.125	.111	.105

TABLE 1—(*Continued*)

r_{pb}	Selection ratio										
	.050	.100	.200	.300	.400	.500	.600	.700	.800	.900	.950
	Proportion of employees considered satisfactory = .20										
.00	.200	.200	.200	.200	.200	.200	.200	.200	.200	.200	.200
.05	.244	.237	.229	.224	.220	.216	.213	.210	.207	.204	.202
.10	.293	.277	.260	.248	.239	.232	.225	.219	.213	.207	.204
.15	.347	.320	.292	.274	.260	.248	.238	.228	.219	.210	.205
.20	.405	.368	.326	.300	.280	.264	.249	.236	.224	.213	.207
.25	.468	.418	.363	.328	.301	.280	.261	.244	.229	.215	.208
.30	.535	.472	.401	.356	.322	.295	.272	.252	.234	.217	.208
.35	.605	.529	.441	.385	.344	.311	.283	.259	.238	.218	.209
.40	.676	.589	.484	.416	.366	.326	.293	.265	.241	.220	.210
.45	.747	.652	.529	.448	.387	.340	.302	.270	.244	.220	.210
.50	.815	.717	.577	.480	.409	.354	.310	.275	.246	.221	.210
.55	.876	.782	.627	.514	.430	.367	.318	.279	.248	.222	.210
.60	.927	.846	.680	.547	.450	.378	.324	.282	.249	.222	.210
.65	.964	.904	.736	.581	.468	.387	.328	.284	.249	.222	.210
.70	.986	.952	.794	.612	.482	.394	.331	.285	.250	.222	.210
.75	.996	.983	.852	.639	.492	.398	.333	.286	.250	.222	.210
.80	1.000	.997	.909	.657	.498	.400	.333	.286	.250	.222	.210
.85	1.000	1.000	.958	.665	.500	.400	.333	.286	.250	.222	.210
.90	1.000	1.000	.990	.667	.500	.400	.333	.286	.250	.222	.210
.95	1.000	1.000	.990	.667	.500	.400	.333	.286	.250	.222	.210
1.00	1.000	1.000	1.000	.667	.500	.400	.333	.286	.250	.222	.211
	Proportion of employees considered satisfactory = .30										
.00	.300	.300	.300	.300	.300	.300	.300	.300	.300	.300	.300
.05	.349	.341	.333	.327	.322	.318	.315	.311	.308	.304	.302
.10	.401	.385	.367	.354	.345	.337	.329	.322	.316	.308	.305
.15	.456	.430	.402	.383	.368	.355	.344	.333	.323	.312	.307
.20	.513	.478	.438	.412	.391	.374	.358	.344	.330	.316	.308
.25	.572	.528	.476	.442	.415	.393	.372	.354	.336	.319	.310
.30	.632	.579	.516	.473	.440	.412	.387	.364	.343	.322	.311
.35	.692	.632	.557	.506	.465	.431	.401	.374	.349	.324	.312
.40	.751	.686	.600	.540	.492	.451	.415	.383	.354	.327	.314
.45	.808	.740	.646	.576	.519	.471	.429	.392	.359	.329	.314
.50	.860	.794	.693	.614	.547	.491	.443	.401	.363	.330	.315
.55	.905	.846	.742	.653	.577	.512	.456	.408	.367	.331	.315
.60	.942	.894	.794	.696	.608	.532	.468	.415	.370	.332	.316
.65	.970	.936	.846	.741	.639	.552	.479	.420	.372	.333	.316
.70	.987	.967	.897	.789	.671	.570	.488	.424	.374	.333	.316
.75	.996	.988	.944	.839	.702	.584	.495	.427	.375	.333	.316
.80	.999	.997	.979	.891	.728	.594	.498	.428	.375	.333	.316
.85	1.000	1.000	.997	.941	.745	.599	.500	.428	.375	.333	.316
.90	1.000	1.000	1.000	.982	.750	.600	.500	.429	.375	.333	.316
.95	1.000	1.000	1.000	.982	.750	.600	.500	.429	.375	.333	.316
1.00	1.000	1.000	1.000	1.000	.750	.600	.500	.429	.375	.333	.316
	Proportion of employees considered satisfactory = .40										
.00	.400	.400	.400	.400	.400	.400	.400	.400	.400	.400	.400
.05	.451	.444	.435	.429	.424	.420	.416	.412	.408	.405	.403
.10	.504	.488	.470	.458	.448	.439	.432	.424	.417	.409	.405
.15	.557	.534	.506	.487	.472	.459	.447	.436	.425	.414	.408
.20	.611	.580	.542	.517	.497	.479	.463	.448	.433	.418	.410

TABLE 1—(*Continued*)

r_{pb}	Selection ratio										
	.050	.100	.200	.300	.400	.500	.600	.700	.800	.900	.950
Proportion of employees considered satisfactory = .40											
.25	.665	.626	.580	.548	.522	.500	.479	.460	.441	.422	.412
.30	.717	.673	.618	.580	.548	.521	.496	.472	.449	.426	.414
.35	.768	.720	.658	.613	.575	.542	.512	.484	.456	.429	.415
.40	.816	.766	.698	.647	.603	.564	.529	.496	.464	.433	.417
.45	.860	.811	.740	.682	.632	.588	.546	.508	.471	.436	.418
.50	.900	.854	.782	.720	.663	.612	.564	.519	.477	.438	.419
.55	.933	.894	.824	.758	.696	.637	.582	.530	.483	.440	.420
.60	.959	.929	.866	.799	.730	.663	.600	.541	.489	.442	.420
.65	.978	.958	.906	.842	.768	.691	.617	.551	.493	.443	.421
.70	.990	.979	.943	.885	.807	.719	.634	.560	.496	.444	.421
.75	.997	.992	.972	.928	.850	.748	.649	.566	.498	.444	.421
.80	.999	.998	.991	.966	.894	.774	.660	.570	.500	.444	.421
.85	1.000	1.000	.998	.992	.939	.792	.665	.571	.500	.444	.421
.90	1.000	1.000	1.000	1.000	.979	.800	.667	.571	.500	.444	.421
.95	1.000	1.000	1.000	1.000	.979	.800	.667	.571	.500	.444	.421
1.00	1.000	1.000	1.000	1.000	1.000	.800	.667	.571	.500	.444	.421
Proportion of employees considered satisfactory = .50											
.00	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500	.500
.05	.552	.544	.535	.529	.524	.520	.516	.512	.509	.505	.503
.10	.602	.588	.570	.558	.548	.540	.532	.525	.518	.510	.505
.15	.653	.631	.605	.587	.573	.560	.549	.538	.526	.515	.508
.20	.702	.674	.640	.617	.598	.581	.565	.550	.535	.519	.511
.25	.748	.716	.676	.647	.623	.602	.582	.563	.544	.524	.513
.30	.793	.757	.712	.678	.649	.623	.599	.576	.553	.529	.515
.35	.834	.797	.747	.709	.676	.646	.617	.590	.562	.533	.518
.40	.872	.835	.783	.741	.703	.669	.636	.603	.571	.537	.520
.45	.905	.871	.818	.773	.732	.693	.655	.617	.580	.541	.521
.50	.933	.903	.853	.807	.762	.718	.675	.631	.588	.545	.523
.55	.956	.932	.886	.840	.793	.745	.696	.646	.597	.548	.524
.60	.974	.956	.918	.875	.826	.773	.718	.660	.604	.551	.525
.65	.986	.975	.946	.908	.861	.804	.740	.675	.612	.553	.526
.70	.994	.988	.969	.940	.896	.836	.764	.689	.617	.554	.526
.75	.998	.995	.986	.968	.932	.872	.788	.700	.622	.555	.526
.80	1.000	.999	.996	.988	.966	.909	.810	.709	.624	.555	.526
.85	1.000	1.000	.999	.998	.990	.947	.827	.713	.625	.556	.526
.90	1.000	1.000	1.000	1.000	.999	.980	.833	.714	.625	.556	.526
.95	1.000	1.000	1.000	1.000	1.000	.999	.833	.714	.625	.556	.526
1.00	1.000	1.000	1.000	1.000	1.000	1.000	.833	.714	.625	.556	.526
Proportion of employees considered satisfactory = .60											
.00	.600	.600	.600	.600	.600	.600	.600	.600	.600	.600	.600
.05	.650	.642	.634	.628	.624	.620	.616	.612	.609	.605	.603
.10	.697	.684	.667	.656	.647	.639	.632	.625	.618	.610	.606
.15	.742	.723	.700	.684	.671	.659	.648	.637	.626	.615	.608
.20	.784	.761	.733	.712	.695	.679	.664	.650	.636	.620	.611
.25	.824	.798	.764	.740	.719	.700	.681	.663	.645	.625	.614
.30	.859	.832	.796	.768	.743	.721	.699	.677	.655	.630	.617
.35	.891	.864	.826	.796	.768	.742	.717	.691	.664	.636	.619
.40	.918	.893	.855	.823	.793	.764	.735	.706	.675	.641	.622
.45	.942	.920	.883	.851	.819	.788	.755	.721	.685	.646	.624

TABLE 1—(*Continued*)

r_{pb}	Selection ratio										
	.050	.100	.200	.300	.400	.500	.600	.700	.800	.900	.950
Proportion of employees considered satisfactory = .60											
.50	.961	.942	.910	.878	.846	.812	.776	.737	.695	.650	.626
.55	.976	.961	.934	.905	.872	.837	.797	.754	.706	.655	.628
.60	.986	.976	.955	.930	.900	.863	.820	.771	.716	.659	.629
.65	.993	.987	.972	.953	.926	.891	.845	.789	.727	.662	.630
.70	.997	.994	.986	.972	.951	.919	.872	.808	.736	.664	.631
.75	.999	.998	.994	.987	.973	.948	.900	.826	.743	.666	.631
.80	1.000	1.000	.998	.996	.990	.974	.930	.843	.748	.666	.632
.85	1.000	1.000	1.000	.999	.998	.992	.960	.854	.750	.667	.632
.90	1.000	1.000	1.000	1.000	1.000	1.000	.986	.857	.750	.667	.632
.95	1.000	1.000	1.000	1.000	1.000	1.000	.999	.857	.750	.667	.632
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.857	.750	.667	.632
Proportion of employees considered satisfactory = .70											
.00	.700	.700	.700	.700	.700	.700	.700	.700	.700	.700	.700
.05	.746	.739	.732	.726	.722	.718	.715	.712	.708	.705	.703
.10	.787	.776	.762	.752	.744	.737	.730	.723	.717	.709	.705
.15	.826	.810	.791	.777	.766	.755	.745	.736	.725	.714	.708
.20	.860	.842	.819	.802	.787	.774	.761	.748	.735	.720	.711
.25	.890	.871	.846	.826	.809	.793	.777	.761	.744	.725	.714
.30	.916	.897	.871	.849	.830	.812	.793	.774	.754	.731	.718
.35	.938	.920	.894	.872	.852	.831	.810	.788	.764	.737	.721
.40	.956	.941	.916	.894	.873	.851	.828	.803	.775	.743	.724
.45	.970	.958	.936	.915	.894	.871	.846	.818	.786	.749	.727
.50	.981	.972	.953	.935	.914	.891	.865	.834	.798	.755	.730
.55	.989	.982	.968	.952	.934	.912	.885	.851	.811	.761	.732
.60	.994	.990	.980	.968	.952	.932	.905	.870	.823	.766	.734
.65	.997	.995	.989	.981	.969	.952	.926	.889	.836	.771	.735
.70	.999	.998	.995	.990	.982	.970	.948	.909	.849	.774	.736
.75	1.000	1.000	1.000	.996	.992	.984	.968	.931	.861	.776	.737
.80	1.000	1.000	1.000	.999	.998	.994	.986	.953	.870	.778	.737
.85	1.000	1.000	1.000	1.000	1.000	.999	.997	.975	.874	.778	.737
.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.992	.875	.778	.737
.95	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.992	.875	.778	.737
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.875	.778	.737
Proportion of employees considered satisfactory = .80											
.00	.800	.800	.800	.800	.800	.800	.800	.800	.800	.800	.800
.05	.839	.833	.827	.823	.819	.816	.813	.810	.807	.804	.802
.10	.872	.864	.852	.844	.838	.832	.826	.821	.815	.809	.805
.15	.901	.890	.876	.865	.856	.848	.840	.832	.823	.813	.808
.20	.926	.914	.898	.885	.874	.864	.853	.843	.832	.819	.811
.25	.946	.934	.917	.904	.891	.880	.867	.855	.841	.824	.814
.30	.962	.951	.935	.921	.908	.895	.882	.867	.850	.830	.818
.35	.974	.965	.950	.937	.924	.911	.896	.879	.860	.836	.821
.40	.983	.976	.964	.952	.939	.926	.910	.892	.871	.843	.825
.45	.990	.984	.975	.964	.953	.940	.925	.906	.882	.850	.829
.50	.994	.991	.984	.975	.966	.954	.939	.920	.894	.857	.832
.55	.997	.995	.990	.984	.977	.967	.953	.934	.907	.865	.836
.60	.999	.998	.995	.991	.985	.978	.966	.949	.920	.872	.838
.65	1.000	1.000	1.000	1.000	.992	.987	.978	.963	.934	.878	.840
.70	1.000	1.000	1.000	1.000	1.000	.994	.988	.977	.948	.884	.841

TABLE 1—(*Continued*)

r_{pb}	Selection ratio										
	.050	.100	.200	.300	.400	.500	.600	.700	.800	.900	.950
Proportion of employees considered satisfactory = .80											
.75	1.000	1.000	1.000	1.000	1.000	1.000	.995	.988	.963	.887	.842
.80	1.000	1.000	1.000	1.000	1.000	1.000	.999	.996	.977	.888	.842
.85	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.990	.889	.842
.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.990	.889	.842
.95	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.889	.842
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.889	.842
Proportion of employees considered satisfactory = .90											
.00	.900	.900	.900	.900	.900	.900	.900	.900	.900	.900	.900
.05	.928	.924	.920	.917	.914	.912	.910	.908	.906	.903	.902
.10	.949	.944	.937	.932	.928	.924	.920	.916	.912	.907	.904
.15	.966	.960	.952	.946	.940	.935	.930	.924	.918	.911	.907
.20	.978	.972	.965	.958	.952	.946	.940	.933	.925	.916	.910
.25	.986	.982	.975	.969	.963	.957	.950	.942	.933	.921	.913
.30	.992	.989	.983	.978	.972	.966	.959	.951	.941	.927	.917
.35	.996	.993	.989	.985	.980	.975	.968	.960	.949	.933	.921
.40	.998	.996	.994	.990	.987	.982	.976	.968	.958	.940	.926
.45	.999	.998	.996	.994	.992	.988	.983	.976	.966	.947	.930
.50	1.000	1.000	1.000	1.000	.995	.993	.989	.984	.974	.955	.935
.55	1.000	1.000	1.000	1.000	1.000	1.000	.994	.990	.982	.962	.939
.60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.994	.989	.970	.943
.65	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.994	.978	.946
.70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.985	.947
.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.991	.947
.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.947
.85	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.947
.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.947
.95	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.947
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.947
Proportion of employees considered satisfactory = .95											
.00	.950	.950	.950	.950	.950	.950	.950	.950	.950	.950	.950
.05	.969	.967	.964	.962	.960	.959	.957	.956	.954	.952	.951
.10	.982	.979	.975	.972	.970	.967	.964	.962	.959	.955	.953
.15	.990	.987	.984	.981	.978	.975	.972	.968	.964	.959	.956
.20	.995	.993	.990	.987	.984	.982	.978	.974	.970	.963	.959
.25	.998	.996	.994	.992	.990	.987	.984	.980	.975	.968	.962
.30	.999	.998	.997	.996	.994	.992	.989	.986	.981	.973	.966
.35	1.000	1.000	1.000	.998	.996	.995	.993	.990	.986	.978	.970
.40	1.000	1.000	1.000	1.000	1.000	.997	.996	.994	.991	.983	.975
.45	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.997	.994	.988	.979
.50	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.997	.992	.984
.55	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.988
.60	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.992
.65	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.995
.70	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998
.75	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999
.80	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
.85	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
.90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
.95	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1.00	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

For the previous example where: SR = .60, BR = .50, point-biserial r = .55, and proportion of satisfactory selectees equaled .696, the proportion in each quadrant is:

$$A = (.696)(.60) = .42$$
$$B = (.60) - (.42) = .18$$
$$C = (1) - (.50 + .18) = .32$$
$$D = (.50) - (.42) = .08$$

With these values, various "hit rates" and other comparisons of particular interest may be made.

REFERENCES

TAYLOR, H. C., & RUSSELL, T. J. The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology*, 1939, **23**, 565–578.

(Received July 13, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 458–465

JOB ANALYSIS METHOD FOR HEALTH-RELATED PROFESSIONS:

A PILOT STUDY OF PHYSICAL THERAPISTS[1]

NEIL S. DUMAS[2] AND JOHN E. MUTHARD

National Center for Health Services Research and Development, United States Public Health Service

Regional Rehabilitation Research Institute, University of Florida

A method for analyzing work of health personnel was devised and applied in a physical therapy service. Procedures for developing the special language for describing the tasks performed by physical therapists and methods for training observers to prepare sequential reports of the ongoing work of staff are presented. Observers were able to reliably report the detailed characteristics of the tasks in a physical therapy service over an extended period of time. Implications of the method are discussed.

Personnel utilization studies of health and rehabilitation services are urgently needed since the gap between the supply and demand of fully qualified persons has not been narrowing. The Health Manpower Report (National Center for Health Statistics, 1968) revealed, for example, that 39% more physical therapists and 23% more physical therapy assistants were needed to meet the demand for services in 1968. In this light, therefore, it appears necessary to plan for the redistribution of the functions and responsibilities of physical therapists and other health professions to achieve maximum efficiency in training and utilizing all levels of personnel including assistants and aides. While the methodology described in the following pages was applied to a single health profession, it should be equally useful in other areas.

Recently established programs for physical therapy assistants and in-service training for aides have been based on professional judgments regarding the relevance of curriculum content to available job description data. An outstanding example of thorough program planning and implementation is described by Fenderson and Larson (1968). Nevertheless, while this program development was carefully planned, the curriculum was based upon available job descriptions and professional judgments. Thus, it lacks the clear objective linkage to the work of the physical therapy service that intensive job analysis can provide. Other studies of the roles and functions of physical therapists have been conducted by McDaniel (1964), Bright (1962), and Anderson, Aldredge, White, and Wroe (1965), but there has been no thorough objective-analysis scheme devised that would provide an accurate description of tasks performed by the staff.

To initiate a program of research on the use of support personnel in the health-related professions, a more sophisticated method for describing the elements of jobs was needed—a job analysis procedure with several particular characteristics: (*a*) The language of the data and technique should be understandable to the professionals in the field being examined. (*b*) The measurements should be reliable and in interval scales insofar as possible.

---

[1] The complete version of the results and instruments will soon be available in the final report of the Social and Rehabilitation Service Project RD-2870 from the National Technical Information Service, United States Department of Commerce, Springfield, Va., at $3.00 for hardcopy and 65¢ for microfiche. Free copies are available from Regional Rehabilitation Research Institute, 901 Lakeshore Towers, Gainesville, Florida 32601 as long as the supply lasts.

This research was supported in part by a Research Grant No. 22-P-55144/4-02 from the Social and Rehabilitation Service, Department of Health, Education, and Welfare, Washington, D. C.

[2] Formerly with the Regional Rehabilitation Research Institute, University of Florida. Requests for reprints should be sent to Neil S. Dumas, Regional Rehabilitation Research Institute, University of Florida, 901 Lakeshore Towers, Gainesville, Florida 32601.

(*c*) Classification of tasks should be completed on-site to minimize the introduction of errors. (*d*) The resultant data should be directly recorded on worksheets specifically designed for direct processing input. (*e*) The results should include not only the proportion of time each task occurred but also a record of the sequence of events. (*f*) The methodology should be compatible with appropriate systems analytic and operations research techniques so that critical decisions made about specific aspects of the job can be simultaneously related to other important data elements.

Standard job analysis procedures such as those developed for personnel selection and evaluation were reviewed and not deemed suitable. The task inventory approach developed by the Personnel Research Laboratory of the United States Air Force (Morsh, Madden, & Christal, 1961) was considered because it has proven useful, reliable, and is designed to permit ready analysis by computers. Although it does provide for determining the relative frequency with which the task is done and the proportion of time that workers contribute to each task as part of their job, it does not provide a description of tasks as they occur sequentially. Thus, it would not enable the data to be used for such purposes as determining the extent to which specific tasks can be separated from the other work of highly trained professionals and reassigned to technicians. Further, it would not enable simulation models to be prepared that would provide administrators with information regarding the impact of various changes on other aspects of the total system.

The series of studies conducted by Bobbitt and her associates (Bobbitt, Gourevitch, Miller, & Jensen, 1969; Bobbitt, Jensen, & Kuehn, 1964) suggested a model for developing a computerized procedure for analyzing the patterns and sequences of tasks in a job. They developed an approach for describing the social interaction of mother-infant monkeys that provided "a relatively exhaustive description of interactive behavior as it occurs, aimed at discovering those patterns and sequences of behavior which seemed to characterize the interaction [p. 112]." Such a description would enable an investigator to develop new hypotheses that could be put to critical and circumscribed tests.

In their study of patient behavior within a hospital environment, Willems and Vineberg (1969) describe an observational procedure that meets several of the criteria for a suitable technique mentioned above. They obtained continuous descriptive protocols for complete days by having observers tape record their observations in everyday language. Subsequently, they were able to reliably classify the observations into six categories with a high (84%) degree of concurrence between two raters. This procedure suggests methods that can meet many of the criteria set but lacks the precision and computer capability required for large scale studies. The pilot study conducted in the University of Florida physical therapy service demonstrated the character of a new method and its efficacy in meeting the criteria set for a sophisticated task analysis system.

METHOD

Instrumentation

To collect data about the work demands within a clinic, a job-analysis tool called the Physical Therapy Observation Schedule (PTOS) was developed. This technique is a highly structured language for the detailed recording of events and is based upon significant tasks and modalities that were identified through a content analysis of selected physical therapy and medical texts. In using the PTOS, a trained observer constructs a series of "sentences" to describe the tasks performed by a clinic staff member.

Encoding a task. For each task (which begins and ends at the observer's judgment), the following data are recorded in sentence-like form:

1. Action—what is being done. Twenty-eight terms were grouped in three categories, that is, people, data, and things (Fine & Heinz, 1958). From the perspective of the observer, a task statement might be "the therapist (I'm observing) is action (doing something). . . ." One term is chosen.

2. Object—type of action in progress. Forty-seven terms were grouped but not fixed in parallel, people, data, and thing categories. These terms are used to expand a description of the action, for example, "the therapist (I'm observing) is *preparing* (action) a bandage (object). . . ." One term is chosen.

3. Person—individuals involved other than staff member being observed. Sixteen categories and combinations of individuals "to, for, or with whom" the action is carried out, for example, "the therapist (I'm observing) is *preparing* (action) a *bandage* (object) for a *patient* (person). . . ." This category enumerates

ACTION (1)

PEOPLE

- 110-ASSISTS
- 121-PREPARES P
- 122-TRANSPORTS P
- 131-HANDLES
- 141-COMMUNICATES
- 142-PARTICIPATES
- 143-OVERSEES
- 150-EVALUATES
- 161-SUPERVISES
- 162-TEACHES
- 163-CONSULTS
- 170-DIRECTS
- 199-OTHER*
- 999-TRAVELS

DATA

- 210-COPIES
- 220-RECORDS
- 230-READS
- 240-WRITES
- 250-PLANS
- 299-OTHER*

THINGS

- 310-PREPARES T
- 311-TRANSPORTS T
- 320-OPERATES
- 330-CONSTRUCTS
- 340-ADJUSTS
- 350-SETS-UP
- 360-APPLIES
- 399-OTHER*

*EXPLANATORY NOTE REQUIRED

OBJECT (2)

- 100-ACTION/TECHNIQUE
- 110-EXERCISES
  - 111-BREATHING
  - 112-MUSCLE RE EDUC.
  - 113-RELAXATION
  - 114-STRENGTH. GEN'L
  - 115-BALANCING
  - 116-AMBULATION
  - 117-DEVELOPMENTAL
  - 118-POSTURE
  - 119-SPEC. STRENGTH
- 120-FIRST AID
- 130-POSITIONING
- 140-BANDAGING
- 150-TRANSPORTATION
- 160-SELF CARE
- 170-PATIENT RESPONSE
- 180-PROFESSIONAL*
- 199-OTHER*

- 200-PROGRESS NOTES
  - 201-INITIAL
  - 202-FINAL
- 210-REPORT
  - 211-STATISTICAL
  - 212-NARRATIVE
- 220-RECORDS
  - 221-ATTENDANCE
  - 222-APPOINTMENT
  - 223-MEDICAL CHART
- 230-MESSAGE
- 240-TREATMENT PROGRAM
- 250-INVENTORY
- 260-EVALUATION FORM
- 270-INSTRUCTIONS
- 299-OTHER*

- 310-EQUIPMENT
- 320-PROSTHESIS
- 330-ADJUNCT
- 340-BANDAGE
- 350-ASSISTIVE DEVICE
  - 351-BRACE
  - 352-CRUTCH
  - 353-WALKER
  - 354-CANE
  - 355-WHEELCHAIR
- 360-PHYSICAL PLANT
- 399-OTHER*

LEVEL (5)

- 1-RPT
- 2-ASST+
- 3-ASST
- 4-AIDE+
- 5-AIDE

PERSON (3)

- 100-PATIENT
- 010-PT-AIDE
- 011-AIDE
- 012-PT-PT
- 013-PT
- 014-STUDENT
- 015-AIDE-AIDE
- 016-ASSISTANT
- 017-PT-ASST
- 018-ASST-ASST
- 019-ASST-AIDE
- 020-PHYSICIAN
- 021-NURSE
- 022-OT
- 023-OTHER HRP
- 030-CLERICAL
  - 031-RECEPTIONIST
  - 032-SECRETARY
- 040-RELATIVE
- 060-ALONE
- 099-OTHER*

MODE (4)

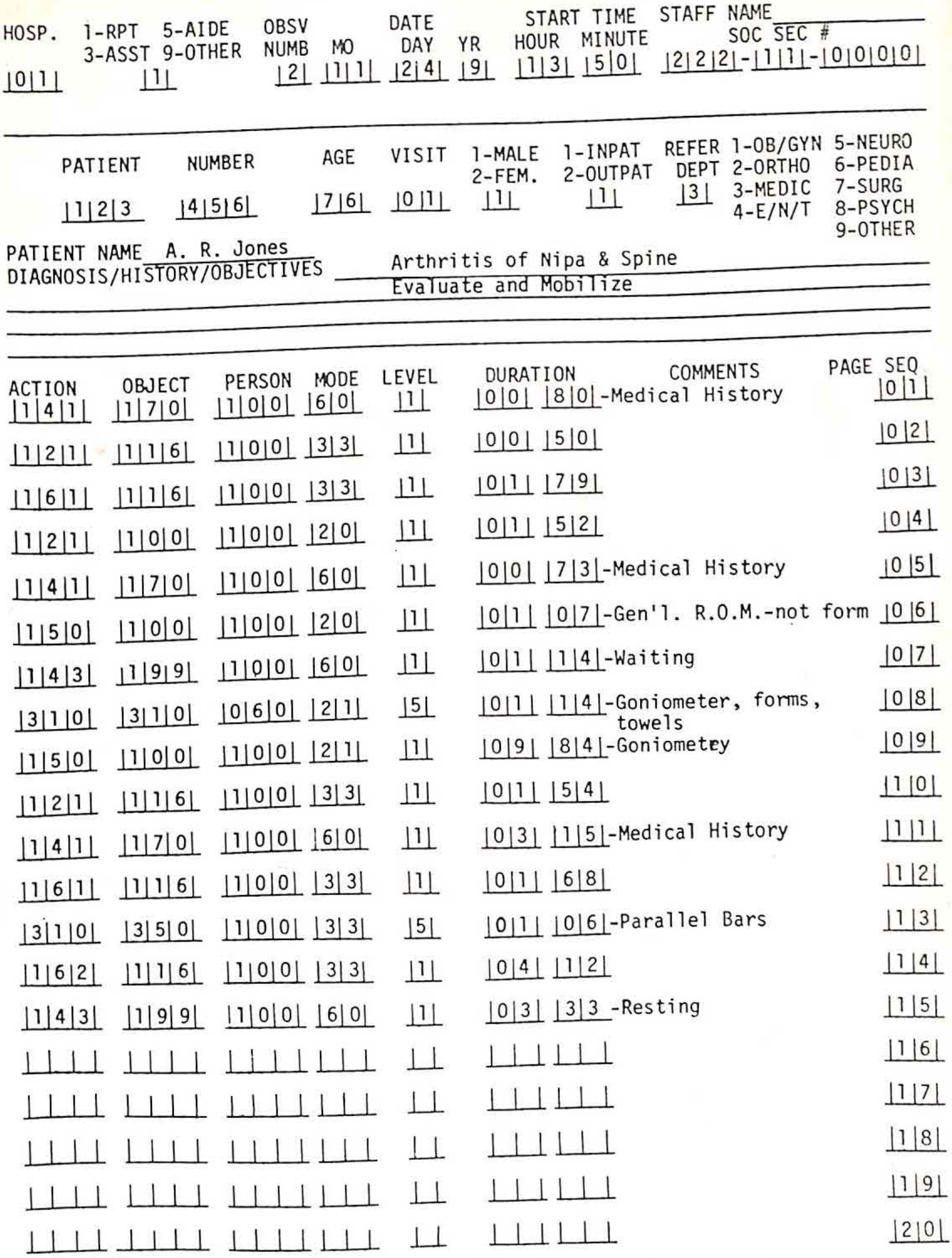
- 01-THERMOTHERAPY
  - 02-CRYOTHERAPY
  - 03-DIATHERMY MW
  - 04-DIATHERMY SW
  - 05-HOT PACK
  - 06-INFRARED
  - 07-PARAFFIN
  - 08-ULTRASOUND
  - 09-ULTRAVIOLET
- 10-HYDROTHERAPY
  - 11-CONTRAST BATH
  - 12-HUBBARD TANK
  - 13-WHIRLPOOL
- 20-EVALUATION
  - 21-ROM
  - 22-M.M.T.
  - 23-ELECTRODIAG.
- 30-INSTRUCTION
  - 31-GEN'L EXER.
  - 32-THER. EXER.
  - 33-GAIT TRAIN.
  - 34-A.D.L.
  - 35-AMPUTEE
- 40-OTHER MEDICAL
  - 41-TRACTION
  - 42-ELEC. STIM.
  - 43-POST. DRAIN.
  - 44-TILT-TABLE
  - 45-MASSAGE
  - 46-PASS. ROM
- 50-CLERICAL*
- 60-PROFESSIONAL*
- 99-OTHER*

FIG. 1. Physical Therapy Observation Schedule coding table.

the individuals involved in the task being observed. This does *not* include the staff member under direct observation who, of course, must always be present. One or more terms are chosen.

4. Mode—Thirty-three basic divisions of treatment or activity. One term is chosen.

5. Level—the staff member's judgment of the complexity of the task just completed. That is, at



HOSP.	1-RPT 5-AIDE 3-ASST 9-OTHER	OBSV NUMB	DATE MO	DATE DAY	DATE YR	START TIME HOUR	START TIME MINUTE	STAFF NAME SOC SEC #
01	1	2	11	24	9	13	50	222-11-0000

PATIENT NUMBER	AGE	VISIT	1-MALE 2-FEM.	1-INPAT 2-OUTPAT	REFER DEPT	1-OB/GYN 2-ORTHO 3-MEDIC 4-E/N/T	5-NEURO 6-PEDIA 7-SURG 8-PSYCH 9-OTHER
123 456	76	01	1	1	3		

PATIENT NAME A. R. Jones

DIAGNOSIS/HISTORY/OBJECTIVES Arthritis of Nipa & Spine

Evaluate and Mobilize

ACTION	OBJECT	PERSON	MODE	LEVEL	DURATION	COMMENTS	PAGE SEQ
141	170	100	60	1	00 80	-Medical History	01
121	116	100	33	1	00 50		02
161	116	100	33	1	01 79		03
121	100	100	20	1	01 52		04
141	170	100	60	1	00 73	-Medical History	05
150	100	100	20	1	01 07	-Gen'l. R.O.M.-not form	06
143	199	100	60	1	01 14	-Waiting	07
310	310	060	21	5	01 14	-Goniometer, forms, towels	08
150	100	100	21	1	09 84	-Goniometey	09
121	116	100	33	1	01 54		10
141	170	100	60	1	03 15	-Medical History	11
161	116	100	33	1	01 68		12
310	350	100	33	5	01 06	-Parallel Bars	13
162	116	100	33	1	04 12		14
143	199	100	60	1	03 33	-Resting	15
							16
							17
							18
							19
							20

Fig. 2. Physical Therapy Observation Schedule coding form.

the completion of each task (i.e., when it is being coded) the observer asks "What is the lowest level of staff member you would have permitted to perform the task you just completed with the patient (if applicable) you just helped. . . ?" This judgment is on a 5-point scale: (1) therapist, (2) experienced assistant, (3) assistant, (4) experienced aide, and (5) aide.

In addition to the above, the duration of each task is recorded in minutes and hundredths of a minute. The recording procedure also indicates the sequence in which it occurred via prenumbered lines. Figure

1 is a copy of the coding table used by the observers as part of their equipment, while Figure 2 displays a completed coding form.

Observer Training

Flash cards and content test. To field test this encoding technique, four licensed physical therapists were recruited and trained as observers. Each observer was initially furnished with a copy of the PTOS Dictionary and a set of flash cards. The Dictionary provides a definition of each of the 129 terms used in the PTOS via narrative or enumeration of as many unique examples of each as possible. The flash cards were used as a training device to aid the observers in memorizing the literal language of the definitions. This proved to be an easy task since the language of the Dictionary originally came from physical therapy and medical texts and was familiar to the observers. A test comprising 81 examples, selected as a stratified sample from the Dictionary, was administered to all four observers. Given 30 min. to select the proper codes for each item (i.e., examples selected from definitions), three observers scored 96% (three wrong) and one scored 100%.

Videotape training. In order to more adequately simulate actual field conditions and improve the reliability of the PTOS in the observers' hands, 11 1-hr. videotapes of actual therapists, aides, and patients were used for training and testing as follows: (*a*) Each observer viewed one tape individually and encoded the tasks seen. (*b*) All observers convened as a group to review the same tape and adjudicate the coding differences until a consensus was derived. (*c*) Steps *a* and *b* above were repeated for eight tapes. (*d*) Each observer individually viewed all eight tapes for a second time in three consecutive sessions and once again coded the events. (*e*) The observer group convened to discuss this latest coding and compared it to the previous group consensus of Steps *b* and *c* above. (*f*) Reliability tests were conducted using the remaining tapes.

Reliability Tests

Reliability measure. With the larger number of possibilities (over 43,000) inherent in the coding scheme, any identical judgments on the part of the observers during a reliability test would be significantly different statistically from chance expectation. A more practical measure, and the one employed here, is to ask whether four trained individuals can watch a continuous stream of events and (*a*) decide that a task is beginning at some point in time, (*b*) decide that a task has ended at some point in time, and (*c*) label or encode the task with the same or similar three-part identifier. Essentially, what was being questioned was whether these four observers could, at any moment in time, identically encode the event in front of them.

The reliability measurement unit consisted of code in force during time cut into discrete units of .10 min. That is, the first minute of the test comprised 10 intervals during which some task or tasks occurred. If one observer reported that a single task occurred during these 10 intervals while a second indicated that another task began in the last interval, there would be 9 intervals of perfect agreement and 1 interval of disagreement (i.e., two observers agreeing 90% of the time). For each interval, we were able to count five possible occurrences: (*a*) four agreeing, (*b*) three agreeing, (*c*) two versus two, (*d*) two versus one versus one, and (*e*) four different codes.

The only modification of this procedure was with respect to transition points, that is, moments at which one task ended and another began. For transition points the time interval was enlarged to a maximum of .30 min. (18 sec.). It was felt that if the observers could make the decision that a task had ended within 18 sec. of each other, then this should be considered close enough to be called simultaneous. Since this methodology was designed to aid in practical decision making, there is little reason to be concerned with events taking less than the time necessary for a more extensively trained professional to call an assistant to do the job for her. With this rationale, accuracy within ±.30 min. (18 sec.) in measuring the duration of a task was judged sufficient. The judgments made by the observers during this larger time unit (maximum of three intervals) were, therefore, recorded as simultaneous and identical for all 18 sec. (if the same code was selected).

Field Test Procedures

Sampling. Four hours of observations (two each in the morning and afternoon) were made for each of 25 consecutive weekdays for a total of 100 hr. of data. An additional 3 days' data were also collected at the start of the study. Without the knowledge of the observers, it was determined prior to data collection that these 3 days were designated as a shakedown period and, as such, considered to be potentially contaminated, that is, by observer nervousness and abnormal staff reactions until the novelty of the presence of a new person in the clinic wore off. The observers, who were licensed physical therapists themselves, wore their uniforms in order to seem as much as possible a natural part of the clinic and minimize this contamination.

Each observer received only her own portion of the sampling schedule and kept this information concealed from the clinic staff for obvious reasons. In order to provide for sickness or other problems that might prevent an observer from locating a particular staff member, a randomization table was provided for the observers' use. Using this guide, an observer was able to randomly select another staff member to follow if the assigned person was unavailable or could not be found. In point of fact, this device was used for 15 out of 50 2-hr. observation periods without upsetting the balance of the sampling plan.

Data collection. Just prior to each starting time, each observer checked her equipment (clip board, stop watches, coding table, and forms, etc.) and located her assigned staff person or used the randomization table to select another *S*. At the scheduled time, the observer began following her assigned *S*, encoding the nature-duration-complexity and sequence of each different task performed by that person during the next 2-hr. period. In addition, the following data were collected once (on the first coding form) for each 2-hr. observation period as a whole: (*a*) hospital identifier, (*b*) classification of person being observed (therapist or aide), (*c*) observer identifier, (*d*) date, (*e*) starting hour and minute, and (*f*) staff name and identifier. A separate coding form was completed for each patient treated during the observation period. In addition to the other identifying information, it included: (*a*) patient identifier, (*b*) age, (*c*) number of previous visits, (*d*) sex, (*e*) inpatient or outpatient status, (*f*) referring department, and (*g*) admission cause–history.

Oversee. Because of its unusual meaning and effect on the data collection, the action oversee deserves a few special comments. Oversee refers to those busy times when a therapist is working with one patient and is simultaneously watching or supervising another. In this case, it is equally unreasonable to claim that the therapist: (*a*) spent 1 hr. with the first patient and no time with the second, (*b*) spent ½ hr. with each, or (*c*) spent 1 hr. with each. To describe this special phenomenon, therefore, the term oversee was devised, since in the above illustration a second staff member would have been required if the therapist had not been able to supervise the second patient.

A single example should sufficiently illustrate the way in which the term oversee was employed. Consider a therapist who has just completed all the steps leading to a whirlpool bath, including positioning of the patient, etc. The therapist now moves on to Patient B to provide some appropriate treatment but keeps an intermittent visual check on Patient A's status. The observer following this therapist would have first entered the oversee code as the last entry on the coding form that has Patient A's identifying information on top. Next, the observer would have started a new coding form for Patient B by filling in his name and hospital identifier (to obtain other required information after the 2-hr. period) and encoding the first task. If, for example, the therapist completes Patient B's treatment and returns to Patient A, then the observer follows the procedure just described as if the latter had not been seen before. In this case, it is not uncommon to find two or three different coding forms completed for the same patient during a single 2-hr. observation period. By comparing the real time difference between the insertion of the oversee code and the therapist's return to that same patient, the duration of this subtle manpower phenomenon can be computed.

RESULTS AND DISCUSSION

Reliability Results

A test of the four observers' ability to reliably encode tasks were performed using previously unseen videotapes of actual and hypothetical patients to simulate a sample of the same audiovisual stimuli that were likely to be encountered in clinics. The initial videotape test comprised eight sequences (47.2 min.) and included 35 tasks. During the test, each observer made four brief judgments and transferred three additional data to encode a single task. These seven items are:

I. Judgments

1. Choose a single code describing the action observed ($N = 28$).
2. Choose a single code describing the object (i.e., type) of action observed ($N = 47$).
3. Choose a single code describing the mode (sic modality) in which the action-object should be classified ($N = 33$). This three part judgment-encoding procedure has $N = 43{,}428$ possibilities.
4. Decide that the task has ended and the next event begun.

II. Data transfer

1. Having decided that a task has ended (I.4 above), record the actual duration in minutes and hundredths of a minute.
2. Encode the people (staff and others) who participated in this task (person).
3. Record the staff member's judgment of the task's level (rating of the minimum level of experience and training the observed staff member thinks is required to have performed the task just completed with the particular patient [if any] and that was just treated). This was not part of the reliability tests.

The results of four observers independently encoding the identical 47.2 min. (472 intervals) of videotape prior to the data collection period appear in Table 1. Entries in the cells are frequencies and cumulative row percentages reading left to right. For example, in 341 of the 472 intervals (72% of the time) all four observers chose the same action code. In 92 of the 472 intervals, three observers chose the same action code. Thus, three or more observers chose the same code 92% of the time.

A second reliability test was performed immediately after the field test was conducted

TABLE 1

OBSERVER AGREEMENT BEFORE DATA COLLECTION

Code category[a]	No. observers agreeing				
	4	3	2–2	2–1–1	None
Action					
No.	341	92	23	9	7
Cumulative %	72	92	97	99	100
Mode					
No.	450	14	8	0	0
Cumulative %	95	98	100	100	100
Action & Mode					
No.	322	99	37	0	14
Cumulative %	68	89	97	97	100
Person					
No.	458	5	0	9	0
Cumulative %	97	98	98	100	100
Transition points					
No.	18	10	5	[b]	[b]
Cumulative %	55	85	100	—	—
Object					
No.	129	116[c]	20	197	10
Cumulative %	27	52	56	98	100
Action–Mode & Object					
No.	115	99[c]	9	223	26
Cumulative %	24	45	47	95	100

[a] The percentages represent the cumulative proportion of the codes for which there was agreement. Thus, the proportion of codes for which there was agreement among three observers also includes the percentage for which there was unanimity.

[b] Nonapplicable.

[c] Usable for explication purposes only.

in the physical therapy clinic at the University of Florida, J. Hillis Miller Health Center. Again, all four observers independently encoded identical videotape lasting 93.6 min. (936 reliability intervals). The results of this second reliability test appear in Table 2.

For the purposes of future use, it was determined on the basis of both reliability tests that the action-mode combination ($N = 924$) would be the best basic unit of observation since it demonstrated a very good reliability (greater than 86% of the time, three or more observers agreeing) both before and after data collection. The observers' ability to enumerate the persons present at any time is almost beyond question (greater than 96% of the time, three or more observers agreeing). As stated, the duration of any identified task is based upon the observers' judgments as to when it begins and when it ends, that is, "transition points." This measure also meets a relatively high standard of reliability with three or more observers agreeing more than 84% of the time.

The weak spot in the PTOS technique appears to be the use of the object codes to explicate the action-mode combinations. The pretest showed a 45% reliability (three or more observers agreeing) while the posttest resulted in a 67% agreement at that level. Since this would mean that the object code in combination with the action-mode unit is likely to be correct only 60% of the time (a weighted average of 45% and 67%), it was decided to reserve its use to descriptive rather than numerical reporting, that is, for explaining action-mode combinations that have high frequencies. If desirable, however, a reliable action-object-mode combination seems feasible if the circumstances are such that the utility of the resultant data would warrant the increased costs for modifying the Dictionary and extending the training of observers.

Field Test Results and Implications

The field test produced some interesting data that bear both on the methodology and

TABLE 2

OBSERVER AGREEMENT AFTER DATA COLLECTION

Code category	No. observers agreeing				
	4	3	2–2	2–1–1	None
Action					
No.	656	225	18	37	0
Cumulative %	70	94	96	100	100
Mode					
No.	562	313	57	2	2
Cumulative %	60	93	99	99	100
Action & Mode					
No.	371	435	42	69	19
Cumulative %	39	86	90	97	100
Person					
No.	890	16	26	4	0
Cumulative %	95	96	99	100	100
Transition points					
No.	36	13	9	[a]	[a]
Cumulative %	62	84	100	—	—
Object					
No.	283	379	110	160	4
Cumulative %	30	70[b]	82	99	100
Action–Mode & Object					
No.	245	384	93	183	31
Cumulative %	26	67[b]	77	96	100

[a] Nonapplicable.

[b] Usable with reservations.

the utility of the information derived from it. Of the 29 action codes, only 28 were used by the observers. Similarly, only two-thirds (22) of the modality codes were employed. In all, 41 different task statements (action-mode combinations) were used to encode more than 80% of the 6,260 events observed over a 25-day period. This suggests that the problem of analyzing the work of the physical therapist relative to patient demands is not only subject to solution but also involves only a limited range of observable phenomena.

Preliminary analysis of the data from this one-case study provided other insights into the clinic operation. While almost one-third of the work now performed by therapists was judged suitable for reallocation to lesser trained personnel, these observations also produced a view of some of the subtleties involved. In very few cases did the therapists believe that a task or set of tasks should be totally delegated to some other level of personnel. Rather, patterns of systematic judgments about treatments and services to particular classes of patients have begun to emerge. In short, there is at least a good basis now for relating the effect of a task-patient interaction to the complexity of the service to be provided.

If these preliminary data hold up in a larger study that is now being planned, the problem will be altered from one of trying to enumerate and understand a large number of tasks to one of working with a limited number of services provided to a wide variety of patients. The planned approach to optimizing the manpower, equipment, physical plant, and training programs relative to patient demands involves the development of a simulation model and stochastic linear program. Briefly, national and local health statistics provide usable forecasts of the probable number of patients in each admission (treatment) category. The job analysis data could translate these forecasts into amounts of manpower at different skill levels that will probably be required to perform particular tasks with particular classes of patients. This information, in turn, can provide not only estimates of the probable quantities of equipment and physical plant required but also can suggest alternative curricula for training programs that maximize the utility of the personnel being educated. Hopefully, the end result will not only be efficient and effective physical therapy services at the lowest possible cost, but also programs of higher education that provide preparation with increased relevance for patient needs.

REFERENCES

ANDERSON, H. E., ALDREDGE, H. P., WHITE, B. C., & WROE, M. C. The roles of the physical therapist: Their importance to the patient. *Journal of the American Physical Therapy Association,* 1965, **45,** 768–772.

BOBBITT, R. A., GOUREVITCH, V. P., MILLER, L. E., & JENSEN, G. D. Dynamics of social interactive behavior: A computerized procedure for analyzing trends, patterns, and sequences. *Psychological Bulletin,* 1969, **71,** 110–121.

BOBBITT, R. A., JENSEN, G. D., & KUEHN, R. E. Development and application of an observational method: A pilot study of the mother-infant relationship in pigtail monkeys. *Journal of Genetic Psychology,* 1964, **105,** 257–274.

BRIGHT, C. P. Professional and non-professional activities of staff physical therapists. *Journal of the American Physical Therapy Association,* 1962, **42,** 769–773.

FENDERSON, D. A., & LARSON, C. W. Planning and establishing a physical therapy assistant program. *Physical Therapy,* 1968, **48,** 963–967.

FINE, S. A., & HEINZ, C. A. The functional occupation classification structure. *Personnel and Guidance Journal,* 1958, 180–192.

MCDANIEL, L. V. The critical incident method in evaluation. *Physical Therapy,* 1964, **44,** 235–242.

MORSH, J. E., MADDEN, J. M., & CHRISTAL, R. E. *Job analysis in the United States Air Force.* (Tech. Rep. No. AD 259 389, WADD-TR-61-113) Lackland Air Force Base, Texas, Personnel Research Laboratory, 1961.

NATIONAL CENTER FOR HEALTH STATISTICS. *Health resources statistics: Health manpower and health facilities.* (Public Health Service Publication No. 1509) United States Department of Health, Education, and Welfare, Washington, D. C., 1968.

WILLEMS, E. P., & VINEBERG, S. E. Direct observations of patients: The interface of environment and behavior. *Psychological Aspects of Disability,* 1969, **16,** 74–88.

(Received May 25, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 466–469

IDENTIFYING PATTERNS OF TEACHER BEHAVIOR FROM STUDENT EVALUATIONS[1]

T. W. FIELD, W. S. SIMPKINS,[2] R. K. BROWNE, AND P. RICH
University of New England, Armidale, Australia

A number of studies have attempted to identify the basic dimensions of variability underlying students' perceptions of teachers. It is possible that these perceptions reflect stylistic differences among teachers. Applying the method of discriminant analysis, it was shown that distinctive patterns of items were characteristic of students' discriminations. These patterns were interpreted as representing important configurations of teaching behavior, and the combination of behaviors for each teacher was viewed as a description of his personal teaching style. The teachers in the study were clearly differentiated in terms of these patterns of teaching behavior.

Student evaluation of teaching is increasingly accepted as a normal part of higher education. It raises questions, however, concerning the basis of student judgments and the extent to which students discriminate between different patterns of teaching behavior.

Some recent studies (Bendig, 1954; Isaacson et al., 1964; Ryans, 1963; Smalzreid & Remmers, 1943) have factor analyzed students' ratings of teachers and described behavioral dimensions on which teachers vary, but the separate factors do not represent specific kinds of teachers. As Morrison and McIntyre (1969) assert "it is the patterns in which they are combined in individuals which characterize their teaching [p. 134]."

In the present study, multiple discriminant analysis was applied to student perceptions of teachers with the purpose of (*a*) identifying patterns of teaching behavior by which students were able to differentiate most clearly among a group of teachers, and (*b*) obtaining the relative standings of teachers on these patterns. It was argued that a teacher's standing indicates the distinctive combination of teaching behaviors that characterize his personal teaching style.

METHOD

Teaching behavior was analyzed by means of an 18-item questionnaire comprising the three highest loading items from each of the six basic factor dimensions in the Isaacson et al. (1964) study. The factor dimensions are described in Table 1.

The questionnaire of 18 items was applied to a group comprising 57 students undertaking an Arts degree at the University of New England, Australia, and who were enrolled in a course of studies presented by five teachers. The students, with at least 1 yr.'s academic experience at the University, were mainly in residence, and the course group was small enough to allow frequent personal contacts among the students and between students and teachers. Thus, they had an opportunity to observe teachers closely and discuss among themselves their teachers and their course experiences. Further, a course was selected that was presented by a panel of teachers. The students enrolled in this course were in a position to make considered judgments of and comparisons between teachers within a recognized subject field.

The questionnaire was completed by each student six times. First, they were asked to describe their conceptions of an ideal teacher in terms of the 18 scale items. Then, on successive uses of the questionnaire, they were asked to describe each of the five teachers who had been responsible for presenting the course. All six sets of responses were included in the analysis, as it was felt that a study of students' perceptions of the ideal together with the actual might provide a broader understanding of the configurations of teaching behavior on which teachers might vary.

In a preliminary analysis, the matrix of intercorrelations for the 18 items was obtained from the pooled sets of responses and then factor analyzed using the principal axis method. The obtained factor loadings were rotated to Kaiser's (1959) Varimax solution, which was used to check the underlying dimensions of the 18 items with those of the original questionnaire.

To determine the extent to which students discriminated the six teachers (including the ideal) the six sets of responses were analyzed using the multiple discriminant analysis technique (Cooley & Lohnes, 1962; Veldman, 1967). The advantage of this method was that it included all 18 items in combination and permitted a comparison of patterns of behavior reflected in item groupings. Correlations between each of the items and the obtained discriminant functions were used to interpret the patterns of teaching behavior described by each discriminant function.

---

[1] This study was supported by a research grant from the University of New England, Australia.

[2] Requests for reprints should be sent to W. S. Simpkins, Faculty of Education, University of New England, Armidale, New South Wales, Australia 2351.

Teacher centroids for each pattern in simulated three-dimensional space were used to indicate the relative standings of the teachers on each pattern.

RESULTS

The factor analysis of the item responses resulted in patterns of loadings very similar to those reported for the original, full questionnaire. In each case, the three items selected to represent each of the original factors loaded highly together. The results showed, in fact, that the original factors adequately described the dimensions of student ratings of teachers as evident in the responses of Australian students to the selected 18 items.

The test for significance of the differences among the six teachers on the 18 items gave rise to a Wilk's lambda of .1835. The F ratio for this discriminant solution was 7.213 with $df = 90, 1552$. The probability of obtaining an F as large as this or larger is very much less than .001, and the hypothesis that these six teachers can be considered as random samples from a common population is rejected.

The discriminant analysis gave rise to five discriminant functions, the latent roots for which are shown in Table 2 together with details of chi-square tests of significance.

While all five roots were significant at better than the .05 level it can be seen that about nine-tenths of the information concerning teacher separation is associated with the first three roots,[3] and since the fourth and fifth roots were more inaccurate, only the first three functions are used in describing the discriminant space. Table 3 shows those correlations of discriminant and original variables, greater than .30, used in interpreting the three-dimensional space.

An examination of the correlations of the original variables with the discriminant functions strongly suggests that each of the three functions defined meaningful patterns of teaching behavior. The configurations of teaching behavior represented by the three major discriminant functions are described below.

TABLE 1

SIX DIMENSIONS OF STUDENT EVALUATION OF TEACHING

Factor	Description (behaviors defining)
Skill	General instructional skills
Overload	Amount and difficulty of work expected by teacher
Structure	Teacher's course organization and planning
Feedback	Teacher's expressed evaluations of student performance
Group interaction	The free interchange between teacher and group
Rapport	Friendly interest on the part of the teacher

Note. See Isaacson et al., 1964.

Discriminant Function 1

This function accounted for the greatest amount of the variance (50.7%) with 12 of the items in the inventory highly correlated with the discriminant function. It appears to describe an interrelated pattern of teaching behavior that draws elements from a number of the basic factors, rapport, interaction, feedback, overload, and structure, the latter being negatively correlated.

A tentative interpretation of the number and types of scale items suggests a configuration of teaching behavior based on the readiness of staff to establish a personal relationship with students based on student participation and, within this form of staff-student interaction, to praise student efforts where

TABLE 2

LATENT ROOTS AND TESTS OF THEIR SIGNIFICANCE

No.	Root	Percentage of trace accounted for[a]	χ^2	df	p
1	1.139	50.7	250.94	22	.001
2	.546	24.3	143.83	20	.001
3	.355	15.7	100.21	18	.001
4	.126	5.6	39.13	16	.01
5	.080	3.5	25.32	14	.05

[a] Trace of $W^{-1}A = 2.2459$ (total discriminating variance).

---

[3] That is, about 91% of the total discriminating variance is accounted for by the first three functions.

TABLE 3
CORRELATIONS OF DISCRIMINANT FUNCTIONS AND ORIGINAL VARIABLES

Factor	Item no.[a]	Correlation with discriminant function		
		1	2	3
Skill	1(19)	—	.76	—
	7(40)	—	.80	—
	13(45)	—	.66	—
Overload	2(8)	.35	.35	—
	8(24)	—	—	—
	14(31)	—	.34	—
Structure	3(15)	−.35	—	—
	9(22)	−.53	—	—
	15(39)	−.41	—	—
Feedback	4(27)	.35	.33	.76
	10(38)	.40	—	.32
	16(44)	—	.39	.44
Group interaction	5(7)	.41	.56	—
	11(10)	.42	.40	—
	17(36)	.53	.52	—
Rapport	6(16)	.53	.40	—
	12(21)	.56	—	—
	18(32)	.63	—	—

[a] Item numbers referred to were those used in the questionnaire in Isaacson et al. (1964).

merited, to set stimulating goals, and to avoid overcontrol of the learning situation.

It calls for comment that these behaviors provided the clearest differentiation between teachers only when viewed in conjunction with each other. Thus, the significance of each element of teaching behavior should not be viewed in isolation from the others in the pattern.

Discriminant Function 2

This function accounted for 24.3% of the variance, with 11 of the items highly correlated with the discriminant function.

Again, a meaningful patterning of items can be discerned in the discriminant function that includes elements from a number of the basic factors: skill, interaction, overload, feedback, and rapport.

A possible interpretation suggests a configuration of teaching behavior based on a skillful and stimulating instructional technique, in association with active student participation in instructional activities and the teacher's readiness to make positive and negative evaluations of student work while setting demanding course goals. As for Discriminant Function 1, the significance of any element in this configuration cannot be seen in isolation from other behaviors in the pattern. For example, the importance of the teacher making both positive and negative assessments of students' work cannot be assessed without reference to the need for a skillful and stimulating teaching technique, positive student participation, etc.

Discriminant Function 3

This function accounted for 15.7% of the variance with three items highly loaded on the function. In contrast to the two previous discriminant functions, all the items were associated with a single basic dimension of student evaluations as identified by factor analysis (i.e., feedback). This pattern is defined by the continuous appraisal of students' work on the part of the teacher and by the clear communication of teacher assessments to the student, whether positive or negative.

The three discriminant functions can be viewed as patterns of teaching behavior that constitute important facets of perceived teaching style, and it has been shown that they provided a highly significant differentiation in students' perceptions among the present group of teachers (including the ideal).

It then becomes important to consider the relative positions of teachers on each of these patterns and to examine the composite standing of each teacher as representing, on the basis of the present data, a distinctive teaching style. For this purpose, the respective teacher centroids were calculated and are listed in Table 4.

As suggested, some indication of the patterns of teaching behavior that distinguished each teacher may be obtained by noting his relative position on each of the three discriminant functions. For example, Table 4 shows that, relative to other teachers, Teacher 3 was high on Discriminant Function 1, low on Discriminant Function 2, and high on Discriminant Function 3. Reference to the patterns of teaching behavior associated with each discriminant function gives a general indication of the orientation of his perceived style of teaching relative to the other teachers.

The data as represented in Table 4 show that the respective teacher centroids are quite widely dispersed in discriminant space. The evidence indicates that students discriminated quite finely between six teachers including both the perceived and the student ideal. A separate discriminant analysis undertaken on the five actual teachers obtained a similar degree of separation.

TABLE 4

TEACHER CENTROIDS

Teacher	Coordinates on discriminant function		
	1	2	3
Ideal	1.25	5.55	2.04
T_1	.08	3.48	2.18
T_2	.55	3.62	1.21
T_3	2.55	3.27	1.90
T_4	2.33	4.03	1.79
T_5	1.46	4.44	.92

DISCUSSION

The considerable degree of differentiation among teachers on each of the patterns suggests that students discriminated quite clearly between teachers in their perceptions.

The patterns of mutually interrelated items indicate that students differentiated between teachers on the basis of distinctive configurations of teaching behaviors. This suggests the possible shortcomings of considering separately various elements and dimensions of teaching behavior. For example, in the second discriminant function, the value of active student participation achieves proper significance only when seen in conjunction with skillful teaching and not as an isolated element of instructional practice.

The relative standings of teachers on the three discriminant functions helped to describe the sets of interrelated behaviors that defined their teaching styles. Also, each discriminant function identified a cluster of interrelated behavioral items that could differ in kind, that is, could be drawn from different dimensions of variability as defined in Isaacson's (1964) study. Discriminant analysis appeared to identify patterns of teaching behavior in student evaluations and thus, merits further consideration as a research tool in the study of teaching styles.

REFERENCES

BENDIG, A. W. A factor analysis of student ratings of psychology instructors on the Purdue scale. *Journal of Educational Psychology*, 1954, **45**, 385–393.

COOLEY, W. W., & LOHNES, P. R. *Multivariate procedures of the behavioral sciences.* New York: Wiley, 1962.

ISAACSON, R. L., et al. Dimensions of student evaluations of teaching. *Journal of Educational Psychology*, 1964, **55**, 344–351.

KAISER, H. F. Computer program for Varimax rotation in factor analysis. *Educational and Psychological Measurement*, 1959, **19**, 413–420.

MORRISON, A., & MCINTYRE, D. *Teachers and teaching.* Harmondsworth, Middlesex, England: Penguin Books, 1969.

RYANS, D. G. Assessment of teacher behaviour and instruction. *Review of Educational Research*, 1963, **33**, 415–441.

SMALZREID, N. T., & REMMERS, H. H. A factor analysis of the Purdue Rating Scale for instructors. *Journal of Educational Psychology*, 1943, **34**, 363–367.

VELDMAN, D. J. *Fortran programming for the behavioral sciences.* New York: Holt, Rinehart & Winston, 1967.

(Received July 8, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 470–476

DIMENSIONS OF VISUAL INFORMATION DISPLAYS [1]

ARTHUR I. SIEGEL [2] AND M. A. FISCHL

Applied Psychological Services, Inc., Wayne, Pennsylvania

Twelve air-defense-oriented visual information displays, representing all combinations of three densities of air threat, two different primary display formats, and two conditions of information coding, were exposed to three *S* groups in order to determine the nature and structure of the display-observer interface. The *S* groups possessed different degrees of familiarity with the displays and the tactical concepts involved in their use. Independent multidimensional scaling analyses were performed on the data derived from each group. Seven dimensions, which were congruent across *S* groups, were isolated. The patterns of factor loadings suggested interpretation in terms of: (*a*) stimulus numerosity, (*b*) primary coding, (*c*) contextual discrimination, (*d*) structure scanning, (*e*) critical relationships, (*f*) cue integration, and (*g*) cognitive processing.

As systems unfold from the conceptual and definition stages to breadboard models, prototypes, and the first production model the engineering psychologist is called on with ever-increasing frequency to provide estimates of the probable effectiveness of alternative design configurations. Quite commonly, the alternative design configurations involve various means of presenting information to the human operators in the system.

Concern for display observers' ways of structuring their percepts has received attention in several recent studies. Silver, Jones, and Landis (1965), in investigating radar training requirements, indicated a pronounced tendency toward stimulus orientation, as opposed to response orientation, on the part of relatively unsophisticated display viewers; viz. "it is clear from the data that the unskilled observers tend to make their judgments in terms of stimulus characteristics rather than in terms of what the situation requires of them [p. 17]." Although framed in a slightly different way, the same concern was expressed by Rigney and DeBow (1966), who directed an investigation into "how the decision-maker arrives at a weighted combination of the different factors upon which an action decision depends [p. 1]." These *E*s investigated the hypothesis that integration of percepts takes place through a process of simplification; that is, selectively ignoring certain characteristics of a multiple-stimulus situation so as to concentrate on only one or two characteristics (Rigney & DeBow, 1966, p. 5). The results of their study seem to have supported the hypothesis tested. Landis et al., (1967) examined experiential backgrounds and display user percepts from a somewhat unique point of view. These *E*s employed *S*s, who characteristically use different perceptual dimensions, to investigate the relationship between multidimensional scaling measures and *S*s' "cognitive-perceptual style [p. 4–7]." Their results indicated that "characteristics of the observer of the display interact significantly with the actual display characteristics in determining the effectiveness of display design [p. 6–1]."

The purpose of the present research was to achieve some level of understanding of the nature of the psychological events that take place when an observer interacts with a display. In addition, the study sought to determine the communalities and differences, if any, that exist across observers of varying backgrounds, orientations, and experiences.

METHOD

The overall method of the present study consisted of: (*a*) development of a set of visual information displays, (*b*) training three *S* groups in the reading and the tactical use of these displays, (*c*) obtaining distance estimates for all possible pairs of displays from each *S* group, (*d*) analyzing the distance estimates of each *S* group by means of standard multidimensional scaling techniques, and (*c*) testing the resulting factors for congruence among groups.

---

[1] This research was supported by the Engineering Psychology Branch, United States Office of Naval Research, under Contract N00014-66-C0183. Appreciation is expressed to James Miller, Marshall Farr, Gerald Malecki, and Martin Tolcott of that agency for encouragement and comments.

[2] Requests for reprints should be sent to Arthur I. Siegel, Director of Applied Psychological Services, Inc., Science Center, Wayne, Pennsylvania 19087.

Visual Displays

The visual information displays employed were static displays of air-defense command-and-control situations. Two major east coast air-defense sectors (Philadelphia and Washington-Baltimore) were represented. The air threats against the two sectors, intelligence information about the threats, relevant meteorological information, and status information concerning the defensive resources were displayed.

In the preparation of these displays, three independent variables were systematically manipulated: threat density (three levels), format (two levels), and color (two levels). The three threat-density levels (number of attacking aircraft) employed were: 5 threats (lowest threat density), 10 threats (intermediate threat density), and 15 threats (highest threat density).

Two completely different display formats were used. One format was completely pictorial. This format presented a map of the area with superimposed range rings, battery locations, and threat and environmental information. Surrounding the geographic presentation were pictorial and graphic summary presentations of related and required defensive information. The alternate display format investigated was the tabular display. Each tabular display presented the identical contents as the pictorial, but presented all information in a digital, alphanumeric fashion by means of a large, fragmented billboard.

The third independent display variable that was manipulated was color application. The same pictorial and tabular displays were prepared in alternative versions, with and without the color coding.

Accordingly, the full set of stimuli consisted of 12 displays, that is, each of the 3 threat densities displayed in both tabular and pictorial format, and each of the 2 formats appearing with and without color application. A more complete description of these visual information displays appears in Siegel and Federman (1967).

Subjects

One *S* group consisted of 17 United States Air Force Reserve Officers of the rank of captain and above who possessed extensive experience in operational command-and-control contexts. A second *S* group consisted of 15 male college students who had previously been trained to a high degree of proficiency in the interpretation and use of these displays. The third *S* group was completely naïve as far as these displays and the concepts involved in their use were concerned. This group consisted of 21 male college students who had never seen these displays prior to the study and who, with one exception, had no prior military experience.

Procedure

Each group was convened and was introduced to the study. Then, four slides, which had been prepared for training purposes and which were similar to those to be employed in the study, were presented and described. The four training slides included: (*a*) the pictorial format in black and white, (*b*) the pictorial format in color, (*c*) the tabular format in black and white, and (*d*) the tabular format in color. All of these were at an intermediate threat-density level. The *E* explained and gave examples of the decisions that could be made on the basis of these displays. Next, experience was given in the type of scaling that *Ss* were required to perform during the experimental data collection. A part of the instructions given at this juncture is presented below:

> You are going to be shown some material on slides, and you will be asked to make judgments about them. The materials will be presented in pairs, and for each pair there is a line in this booklet that looks like this:

	Very Similar			Moderately Similar	
	1	2	3	4	5
Comparison #	___	___	___	___	___

Moderately Different				Very Different	
6	7	8	9	10	11
___	___	___	___	___	___

> As you can see, this is an 11-point scale ranging from *1* for "Very Similar," through less and less similarity, to *11* for "Very Different." For each pair of materials that is shown to you, you are to put a check at the place on the scale which comes closest to representing *your judgment* of the materials. REMEMBER, THERE ARE NO RIGHT OR WRONG ANSWERS—THE BEST ANSWER IS THE JUDGMENT OF HOW THE MATERIALS SEEM TO YOU.

The formal data collection was conducted in the same way as the practice scaling except that there were more stimuli (12 instead of 4), and in order to obtain estimates of intrajudge reliability, stimuli were pair-compared twice. The procedure was identical in each of the three *S* groups. The pairs were presented by projecting 1 of the 12 displays as a quasi-standard stimulus against which each of the other 11 was successively compared in terms of similarity. A second display then served as quasi-standard for comparison against each of the remaining 11, and so forth until every display had served once as a standard and 11 times as a comparator. In order to minimize the risk of inadvertently forcing the emergence of one of the independent variables as a dimension, and to control for position and sequence effects, the following controls were instituted:

1. With respect to the three independent variables, the order of appearance of each of the 12 displays as a quasi-standard was: (*a*) simple alternation of low through high density; (*b*) random choice of tabular

TABLE 1

Displays with Heaviest Loadings on Each Derived Dimension

Dimension 1		Dimension 2		Dimension 3		Dimension 4		Dimension 5		Dimension 6		Dimension 7	
Display	Load	Display	Load	Display	Load	Display	Load	Display	Load	Display	Load	Display	Load
Air Force *Ss*													
5-PC	5.38	15-TC	4.40	15-PC	4.38	5-TC	5.24	10-PC	4.99	10-TBW	4.88	10-TC	4.50
5-PBW	4.22	15-TBW	3.36	15-PBW	3.27	5-TBW	3.27	10-PBW	3.06	10-TC	2.10	10-TBW	2.37
10-TC	−.99	5-TBW	1.04	5-TC	−1.92	15-TBW	−1.05	15-PBW	−1.05	5-TBW	−1.43	15-PC	−1.03
10-TBW	−1.35	10-PC	−1.45	5-TBW	−4.24	10-PC	−1.06	5-TC	−1.29	5-PC	−1.52	15-PBW	−1.48
15-PBW	−1.39	5-PBW	−2.89			15-PC	−1.12	5-PBW	−1.42	15-PBW	−1.57	5-PC	−1.65
15-TC	−1.72	10-PBW	−3.60			10-PBW	−1.84	15-TBW	−2.10	15-PC	−2.11	5-PBW	−2.39
15-TBW	−3.14					15-PBW	−2.25	15-TC	−2.38				
Trained college student *Ss*													
5-PC	5.14	15-PC	4.12	10-TC	4.56	10-PC	4.83	5-TC	4.80				
5-PBW	4.97	15-PBW	3.48	10-TBW	4.43	10-PBW	3.59	5-TBW	2.15				
5-TBW	1.12	15-TC	2.63	10-PC	1.23	10-TBW	.98	15-TC	1.02				
10-TBW	−1.25	5-PC	−1.06	5-TBW	−1.09	15-PBW	−1.45	15-TBW	−1.33				
15-PC	−1.64	10-TBW	−1.40	15-TBW	−1.36	15-TBW	−3.22	15-PBW	−1.65				
15-TC	−3.76	5-PBW	−1.70	5-PBW	−1.61	15-TC	−3.27	10-TBW	−1.73				
15-TBW	−4.08	5-TC	−2.23	15-PBW	−2.86			10-PBW	−3.73				
		5-TBW	−4.40	15-PC	−2.97								
Naïve college student *Ss*													
5-PC	5.17	5-TC	4.92	10-TBW	4.54	10-PC	4.17	15-TC	3.93	15-PBW	4.23		
5-PBW	5.10	5-TBW	4.49	10-TC	3.31	10-PBW	2.77	15-TBW	2.04	15-PC	1.72		
10-TC	−1.40	10-PBW	−1.09	15-PBW	1.44	15-TC	−1.57	5-TC	1.62	10-PBW	.98		
10-TBW	−1.50	10-PC	−1.67	5-PC	−1.46	5-TBW	−2.16	10-PC	−2.14	5-TBW	−1.12		
15-TBW	−3.01	15-PC	−2.67	15-PC	−3.83	15-TBW	−3.74	10-PBW	−4.06	5-TC	−1.20		
15-TC	−3.10	15-PBW	−2.74							10-TBW	−1.46		
										10-TC	−3.54		

Note.—Loadings are derived from original data that were distance judgments. Some readers may prefer the alternative term projections to our term loadings. Under display, the number describes threat density. Abbreviations: P = pictorial format, T = tabular format, C = color, BW = black and white.

format to begin, with subsequent simple alternation of the two formats; and (*c*) random choice of color display to begin, with subsequent ABBA sequence.

2. Coin toss determined that the quasi-standard stimulus would appear as the left-hand member of all pairs for the first 66 comparisons and as the right-hand member for the last 66 comparisons.

3. For each of the first six sets of comparisons (11 comparisons per set), a different random sequence of the 11-comparison stimuli was employed. These six sequences were exactly reversed for Sets 7–12, with the quasi-standard stimulus from the corresponding initial set occupying the position vacated by the new quasi-standard stimulus.

The mean and standard deviation of the distribution of similarity judgments of each pair of displays was calculated in each *S* group. The former of these statistics gave the interstimulus distances between the pairs of displays for insertion into the multidimensional analysis; the latter statistic afforded appraisal of the homogeneity of each of the groups' judgments. In addition, since each pair of displays was judged twice by each *S*, the data for an intra-*S* reliability estimate were available. A different one-third of the 66 display pairs was randomly selected for investigation in each *S* group, and the correlation between first judgment and second judgment of these pairs was obtained over all *S*s in the particular groups.

Three independent multidimensional analyses of the interstimulus distances were performed, one for each *S* group. Each analysis began with the original interstimulus distances for the group concerned. The first step was estimation of the "additive constant" (Torgerson, 1958, pp. 268ff) required for converting the obtained comparative distances to absolute distances. Here, the procedure employed was the general solution described by Messick and Abelson (1956). The ensuing matrix of absolute distances was then employed in deriving the matrix of scalar products, which was then factored by the method of principal components with rotation by the Equamax approximation to orthogonal simple structure. The end result was three multidimensional scaling analyses, each representing the structure of the percepts of an individual *S* group.

RESULTS

Intra-*S* reliability was given by the correlation between *S*'s first judgment of a pair of displays and his second judgment of that pair.[3] The resulting standard deviations were of modest size, suggesting reasonable inter-*S* agreement. The median of this set of 66 correlation coefficients was between .62 and .63, which, with *n*s of the sizes involved, is significantly greater than zero ($p < .01$). The *S* groups' experience and background were apparently major influences on the reliability of the judgment process. Of the 33 reliability estimates larger than the median, 17 were derived from the Air Force *S* group, 11 from the trained college student group, and only 5 emanated in the naïve college student group.

The additive constants required to convert the comparative distances (the mean similarity estimates) into absolute distances were 3.09 for the Air Force *S* group, 2.66 for the trained college student group, and 2.51 for the naïve college student group. Each mean was accordingly increased by the magnitude of its respective constant. Factor analysis of the matrices of scalar products derived from the matrices of interstimulus distance judgments, with rotation to orthogonal simple structure, yielded seven dimensions for the Air Force *S*s' percepts, five dimensions for the trained college students' percepts, and six dimensions for the naïve college students' percepts. Displays loaded most heavily on the respective dimensions are shown in Table 1. Three criteria were invoked in order to judge the adequacy of each solution: (*a*) sharpness of break in Eigenvalues, (*b*) amount of predictable variance encompassed, and (*c*) apparent freedom from over- or underspecification.

Examination of Table 1 indicates a high degree of apparent similarity, across *S* groups, in the loading of a number of the dimensions. For example, the first dimension in each *S* group appears to be similarly loaded, the second dimension in the Air Force group appears to be similar to the fifth dimension in the naïve college student group, and the forth dimension in each of the naïve and trained college student groups appears to be similar to the fifth dimension in the Air Force *S* group. To assess these and other apparent similarities, tests of factorial congruence (Harman, 1960, pp. 256–259) were performed. The results of these tests are shown in Table 2. This table shows seven sets of congruence coefficients. Each set, except the second and the last, presents congruence coefficients across all three *S* groups. For the second set there was no dimension in the trained college student group suffi-

---

[3] The mean similarity judgments for each pair of displays, along with the respective standard deviations have been deposited with the National Auxiliary Publications Service. Order Document No. 01473 from the National Auxiliary Publications Service of the American Society for Information Science, c/o CCM Information Sciences, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche and make checks payable to: Research and Microfilm Publications, Inc.

TABLE 2

CONGRUENCE OF DISPLAY DIMENSIONS ACROSS THREE SUBJECT GROUPS

Set	*S* group	Dimension no.	*S* group	Dimension no.	Congruence coefficient
1	Air Force *S*s	1	Trained college students	1	.92
	Air Force *S*s	1	Naïve college students	1	.95
	Trained college students	1	Naïve college students	1	.94
2	Air Force *S*s	2	Naïve college students	5	.84
3	Air Force *S*s	3	Trained college students	2	.90
	Air Force *S*s	3	Naïve college students	6	.68
	Trained college students	2	Naïve college students	6	.55
4	Air Force *S*s	4	Trained college students	5	.88
	Air Force *S*s	4	Naïve college students	2	.93
	Trained college students	5	Naïve college students	2	.77
5	Air Force *S*s	5	Trained college students	4	.92
	Air Force *S*s	5	Naïve college students	4	.88
	Trained college students	4	Naïve college students	4	.92
6	Air Force *S*s	6	Trained college students	3	.86
	Air Force *S*s	6	Naïve college students	3	.88
	Trained college students	3	Naïve college students	3	.91
7	Air Force *S*s	7	Trained college students	3	.88
	Air Force *S*s	7	Naïve college students	3	.80

ciently similar in appearance with the pair shown to warrant quantitative congruence investigation. The last set is merely an elaboration of the sixth set, with Dimension 3 of both college groups being compared for congruence against a different dimension of the Air Force *S* group. Thus, the congruence coefficient across Dimension 3 of the college student groups was not repeated in Table 2.

DISCUSSION AND INTERPRETATION

The seven sets of dimensions reconcile, to greater or lesser degree, all 18 of the factored dimensions obtained from the three *S* groups. That is, the seven dimensions of the Air Force *S* group's percepts, the five dimensions of the trained college students' percepts, and the six dimensions of the naïve college students' percepts all seem to be manifestations of the same common core consisting of seven dimensions of the display-observer interface.

Dimension 1 was the first dimension to emerge from the analysis of the data of each of the three *S* groups. In all three groups, the displays with the heaviest positive loadings on this dimension were the low-threat-density pictorial displays. The displays with the heaviest negative loadings were the high-threat-density tabular displays. While the density bipolarity has interpretative appeal, another pattern exists. With respect to the way the two formats present information, it is apparent that to present the same information the tabular format requires considerably more stimuli than the pictorial. Extending this thinking about number of stimuli, it follows that: (*a*) the fewest stimulus elements should be found in the low-threat-density pictorial display; the largest number of stimulus elements should be found in the high-density tabular display; (*b*) the second fewest stimulus elements should be found in either the low-threat-density tabular display or the intermediate-density pictorial display; and (*c*) rising toward the larger number of stimulus elements should be the high-threat-density pictorial display and the intermediate-density tabular display. This was virtually the exact pattern of loadings on Dimension 1. The dimension, accordingly, was interpreted to reflect *stimulus numerosity*.

Dimension 2 emerged as the second dimension in the Air Force *S* group and the fifth dimension in the naïve college student group. Within these dimensions, the displays with heaviest positive loadings were the tabular displays of both high- and low-threat density. The displays with the heaviest negative loadings were pictorial displays, primarily of intermediate density. This interactive pattern is suggestive of a process in which a display viewer, in seeking to organize the perceptual field, scans for the emergence of a structure. The structure may be given in part by the literal stimuli he sees, but the structure is also given by the meaning he attaches to those stimuli. Accordingly, the dimension was interpreted to represent *structure scanning*, which is believed to be partly perceptual and partly cognitive, and to be manifest during early display involvement but to dissipate with experience with a display.

Dimension 3 appeared third in the Air Force *S* group, second in the trained college student group, and sixth in the naïve college student group. This dimension appears cognitive in nature and to be associated with the number of mental acts to be performed with the displayed information. The evidence for interpreting this dimension as being involved with cognitive activity is drawn from the near-perfect bipolar assignment of the extreme threat densities. For all three groups, displays of high-threat density possessed maximal positive loadings on this dimension, irrespective of color coding or, in the trained group, display format. For the trained college student and the Air Force *S* groups, displays of lowest threat density yielded maximal negative loadings on this dimension, again irrespective of color coding or display format. Thus, over all three *S* groups, the two groups with the greatest familiarity with the displays rendered a perfect density dichotomy. The available evidence seems disposed toward the increase in the number of cognitive acts associated with the comprehension of threat density as the principal process involved. Accordingly, the dimension was interpreted as one of *cognitive processing activity*.

Dimension 4, somewhat like Dimension 2, seems to be associated with the display viewer's organization of his perceptuocognitive field. Unlike Dimension 2, however, the apparent organization seems to be a more sophisticated one involving relationships among a larger number of variables. This dimension emerged as the fourth in the Air Force *S* group, the fifth in trained college student group, and the second in the naïve college student group. With the exception that the low-threat-density tabular displays, in both color and black and white, carry the maximum positive loadings in all three groups, the remainder of the factor loading patterns appear, on the surface, to involve all combinations of density, format, color application, and numerosity. Actually, there seems to be an underlying ordering of these variables, namely, the display viewer's identification of what he perceives as critical relationships. The display viewers apparently simultaneously dealt with critical relationships among numerous perceptuocognitive variables. The particular variables of choice seem a function of observer experience with the displays and the information content, but the organizing probe for critical relationships seems present regardless. Accordingly, Dimension 4 has been named *critical relationships*.

Dimension 5 emerged as the fifth dimension in the Air Force *S* group and as the fourth dimension in each of the student groups. Displays with the heaviest positive loadings on this dimension were pictorial displays of intermediate threat density, with the addition of a tabular display of intermediate threat density in the trained *S* group. The pattern of negative loadings was consistent in the extremes for tabular displays, and reasonably consistent in the extremes for high density. Displays with heavy, but less than extreme, negative loadings showed ordering based on both density and format. Several of the previous dimensions also seemed characterized by an interaction of constructs associated with the displays. One important interpretative difference between the present dimension and any of the others, however, seems to be in the relative weight assigned to cognition in the interactions. Dimension 2 involved a Format × Content interaction, but seemed to be primarily a perceptual dimension. Dimension 3 seemed exclusively cognitive. Dimension 4 involved the interaction of all variables but seemed more heavily weighted toward the sensory and the perceptual activities of relating numerosity (sensory), color (sensory/perceptual), and format (perceptual) to each other and to the one cognitive variable, density. In Dimension 5, the emphasis seems to be on a cognitive integration of cues and their meanings; the dimension was interpreted as *cue integration*.

Dimension 6 emerged as the sixth dimension for the Air Force group and as the third dimension for the other groups. There was also congruence between the Air Force *S* Dimension 7 and Dimension 3 in each of the student groups. To which of the two Air Force *S* dimensions the student Dimension 3 corresponds remains unknown. Inasmuch as the lowest congruence coefficient (.80) in this set of five involved Air Force *S* Dimension 7, Air Force *S* Dimension 7 has been tentatively held out as a unique dimension. Air Force *S* Dimension 6 was interpreted as congruent with the student Dimension 3. Display Dimension 6 bears heavy positive loadings from displays of intermediate threat density, primarily tabular in format, and relatively high stimulus numerosity. Heavy negative loadings are contributed principally by pictorial displays of lower stimulus numerosity. The dimension seems best described as a Numerosity × Format interactive one, involving perceptual discriminations on the basis of physical stimulus contexts. To reflect this, the dimension was named *contextual discrimination*.

Display Dimension 7 bears heavy positive loadings from tabular displays and heavy negative loadings from pictorial displays. Although there appear to be some density groupings, the rationale is not clear. Accordingly, it was interpreted as reflecting the two primary formats in which the sets of displays were encoded and named *primary coding*.

This dimensional structure, along with the interpretation suggested by the structure, seems to possess a number of implications for both the human factors practitioner and the display designer. First, although others have found in other contexts (e.g., Fleishman, 1954) that a factorial structure will vary as the experience level of the persons sampled varies, such longitudinal and developmental effects have not generally been considered at display-observer interfaces. The finding tends to suggest, however, that the optimal display design will vary with the experience level of the user. Thus, the display which is optimal for the user possessing one level of experience may not be optimal for a user who possesses a greater or lesser experience level. Second, we note that many of the dimensions seemed to be based on an interaction of constructs. Dimensions 2, 4, 5, and 6 seemed to be based on a rather complex interaction between display variables and display-observer variables. These findings suggest that typical human engineering checklists, which consider hardware/display variables in a rather simplistic cue manner, can at best, account for only a small part of the variance at the display-observer interface. Such checklists seem to be most concerned with our primary coding factor and seem to have little or no concern with many of our other factors. Extending such checklists to the other constructs indicated by the present analysis may be indicated. Finally, we note that the dimensions seem to fall into three logical groups. *Stimulus numerosity* and *primary coding* seem largely display-characteristic-oriented. *Contextual discrimination*, *structure scanning*, and *critical relationships* seem perceptual in nature, while *cue integration* and *cognitive processing* appear largely cognitive in nature. This grouping may help to define the nature of the events that occur in the display reading–information-processing situation and to provide insight into variables underlying effective display design.

REFERENCES

FLEISHMAN, E. A., & HEMPEL, W. E., JR. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, 19, 239–252.

HARMAN, H. H. *Modern factor analysis*. Chicago: University of Chicago Press, 1960.

LANDIS, D., SLIVKA, R. M., JONES, J. M., et al. *Evaluation of large scale visual displays*. (USAF Tech. Rep. No. RADCTR-67-57) Griffiss Air Force Base, Rome, N. Y.: Rome Air Development Center, 1967.

MESSICK, S. J., & ABELSON, R. P. The additive constant problem in multidimensional scaling. *Psychometrika*, 1956, 21, 1–15.

RIGNEY, J. W., & DEBOW, C. H. *Decision strategies in AAW: I. Analysis of air threat judgments and weapons assignments*. Los Angeles: University of Southern California, 1966.

SIEGEL, A. I., & FEDERMAN, P. J. *Validation of the DEI technique for large scale display evaluation*. Wayne, Pa.: Applied Psychological Services, 1967.

SILVER, C. A., JONES, J. M., & LANDIS, D. *Analysis of radar training requirements*. (NAVTRADEVCEN Tech. Rep. No. 1345-1) Port Washington, N. Y.: Naval Training Device Center, 1965.

TORGERSON, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

(Received April 8, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 477–483

RECOGNITION OF HUMAN FACES:

EFFECTS OF TARGET EXPOSURE TIME, TARGET POSITION, POSE POSITION, AND TYPE OF PHOTOGRAPH[1]

KENNETH R. LAUGHERY,[2] JUDITH F. ALEXANDER, AND ALAN B. LANE
State University of New York at Buffalo

Three experiments investigated *S*s' ability to recognize a target person (whom they had previously seen) in a test series of 150 pictures of faces. The longer the original exposure to the target and the earlier the target picture appeared in the test series, the greater the probability of recognition. The pose position in the test series (front, profile, or portrait view) and the type of photograph (color or black and white) did not affect recognition. These results have implications for procedures used by law enforcement agencies in the process of a witness attempting to identify a criminal from a set of mug shots.

The last 10 yr. have witnessed a tremendous upsurge in psychological research on human memory. More recently, a great deal of this research has focused on visual memory or imagery. A general finding is that when people do form visual images of things to be remembered, retention is remarkably better than when images are not formed (Bower, 1970; Paivio, 1969).

Several visual memory studies have used experimental paradigms that resemble the situation where a witness who has seen a criminal is trying to identify that person's picture from a set of alternatives. Shepard (1967) reported an experiment in which *S*s looked through approximately 600 pictures of "things" and "scenes." The *S*s were then presented with pairs of pictures where one of each pair had been in the original set of 600, and the other had not previously been seen. Each *S* saw 60 such pairs and his task was to note which item in each pair had been seen before. The overall performance was 97% correct. Hochberg and Galper (1967) performed a similar study using pictures of female faces, (taken from a college yearbook). Their *S*s looked at 35 or 60 faces and then 15 pairs. Although recognition was better with 35 faces in the original set, performance was high (about 90%), even with the larger set. In still a third study with this technique Yin (1969) reported 96% correct identifications with pictures of adult males.

In another experiment using a slightly different procedure, Nickerson (1965) showed *S*s 200 pictures of "things." In a second sequence of 400 pictures, *S*s had to say whether each picture had appeared in the original 200. Nickerson systematically varied the lag (number of intervening pictures) between the first occurrence (in the 200 set) and the second occurrence (in the 400 set). With lags of 40 and 200, performance was 97% and 87% correct, respectively.

From the results of these studies one might expect witnesses to be extremely good at identifying criminals from a set of mug shots. Yet, this is clearly not the case; at least not at the 90% and higher levels reported in the aforementioned experiments. Of course, there are many reasons why the identification of criminals from law enforcement files would not be expected to be at this high level. The experimental situations described above represent ideal conditions by comparison to the real-world situation. The trauma of being present when a crime is committed, differences between the criminal's appearance during the crime and in the mug file (disguises, out-of-date photographs, etc.), and the fact that a witness may

---

[1] This research was supported in part by Research Grant No. MH-11595 from the National Institute of Mental Health, United States Public Health Service, and, in part, by Research Grant No. 199 from the Law Enforcement Assistance Administration.

The assistance of David Lenorovitz and Gilbert J. Harris in collecting the film library and analyzing data is gratefully acknowledged. The authors are also indebted to Albert Zavala for his valuable comments on an earlier draft of this article.

[2] Requests for reprints should be sent to Kenneth R. Laughery, State University of New York at Buffalo, Department of Psychology, Buffalo, New York 14214.

indeed look through thousands of photographs before seeing the criminal's picture are only a few of the factors that could reduce the probability of a correct identification. On the other hand, there is little in the way of experimental evidence to indicate just what are the human's abilities and limitations in identifying people from a set of photographic records.

With these considerations in mind, three experiments were designed to explore the effects of several variables on the identification task. The procedure in these studies was intended to resemble the task of a witness attempting to identify a criminal from a set of mug shots. All three studies were essentially exploratory in that they represented an attempt to isolate some of the important factors that determine identification performance.

EXPERIMENT I

The first experiment was undertaken to determine the effects of three task variables on the ability to recognize a face. The variables were target exposure time, target position, and observation instructions.

Method

Subjects. The *S*s were 128 students from an introductory psychology course at the State University of New York at Buffalo. There were 53 males and 75 females. Participation in the experiment partially fulfilled a course requirement.

Task. After viewing four sequentially projected slides showing different candid positions of a particular human face, hereafter referred to as the target, *S*s viewed a test series of 150 projected slides of human faces. The *S*s' task was to indicate on an answer sheet whether each slide in the test series was or was not the target. In addition, *S*s indicated a confidence level in their (Yes/No) decision on a 3-point scale—possibly, probably, or certain.

Design. The design of the experiment was a $2 \times 2 \times 2 \times 2$ factorial with all factors manipulated as between-*S* variables. There were eight *S*s in each of the 16 treatment groups. The variables were manipulated as follows: observation instructions included prior knowledge of the task or no knowledge of the task; target exposure time was 10 sec. or 32 sec. (viewing the four candid pictures for 2.5 sec. or 8.0 sec. each); target position in the 150-item test series was either 40 or 140; and two targets—a Caucasian male with fair colored hair and complexion wearing glasses, or a Caucasian male with darker colored hair and complexion without glasses.

Apparatus. Instructions were presented with a Sony TC 102 tape recorder. Slides were presented with a Kodak Carousel AV 900 projector with a 4–6-in., F3.5 Zoom Ektamar lens onto a Da-Lite screen. For the four initial candid-position slides, the projector was operated manually. During the test series, the projector was set to change automatically at 5.0-sec. intervals. Actually, in the test series every other slot in the slide trays was empty. With the automatic 5.0-sec. change this procedure resulted in each slide being projected on the screen for 4.0 sec. with 6.0 sec. between slides—during which *S*s recorded their responses on the answer sheets.

Materials. The slides were half frame, 35-mm. Ektachrome transparencies showing front bust views of Caucasian males within an age range of 17–23 yr. The four candid-position slides were of this same type except, of course, that instead of a front view, they were candid positions and ranged from left-side to right-side views. The candid positions were obtained by taking a large number of pictures of the target in casual positions and selecting those that seemed least posed.

The test series consisted of 149 decoys and the target, all front bust views. The decoys were randomly selected from a film library of 544 different faces. The physical parameters of all slides were constant (sharpness, scale, lighting, etc.). Faces of fair and darker coloring and with and without glasses were approximately equal in number and randomly distributed throughout the series. The target shown with glasses in the four candid-position slides also wore glasses in the test series, and the target without glasses appeared in the test series without them. The slides were projected so as to be life-size on the screen.

Procedure. Each *S* was given an answer sheet consisting of four pages with two 25-space columns on each page. These spaces were numbered consecutively from 1–200. Each space on the answer sheets corresponded to the one of the test series slides. After viewing a slide, *S* marked either yes or no and indicated a level of confidence (possibly, probably, or certain) in the decision. Although *S*s viewed only 150 slides in the test series, the answer sheet was numbered to 200 in order to eliminate any bias that might result from anticipating the end of the series. The instructions, incidentally, indicated that the target's picture might appear in the test series several times, only once, or not at all.

At the conclusion of the test series, each *S* was given a protocol sheet on which he was to record any particular technique used in making his decisions and the facial features that were most important in his decisions.

The eight *S*s in each of the 16 experimental conditions were run as a group; the groups of *S*s were randomly assigned to the conditions. Four *S*s were seated at each of two long tables, one behind the other, in a normal-sized classroom. Dividers were placed on the tables between the *S*s. The screen was located in the front center of the room at a height slightly above the heads of the seated *S*s. The tables were 8.0 and 14.0 ft. from the screen. The projector was located at the rear center of the room, and the projected light extended across the center of the two tables onto the screen. The room was darkened to assure good vision of the slides but with sufficient light to read and mark the answer sheet.

The first part of the instructions was played to *S*s. In the prior-knowledge conditions the instructions make it clear that *S*s would later be looking for a picture of the person whom they were about to see (in the four candid-position slides). In the no-knowledge conditions, *S*s were simply told that they would be looking at slides containing faces and that the following four slides (the candid-position slides) were examples. If an *S* knew the target, he was (to be) excused from the experiment.

The four candid-position slides were shown for a total of 10 or 32 sec., depending on the experimental condition. Next, the answer sheets were passed out, and the remainder of the instructions, which was the same for all *S*s, was played. Two sample slides were projected at the automatic setting of 5.0 sec. to provide an example of the task and the pacing. The test series was then begun. The interval of time from the end of the observation period to the beginning of the test series was approximately 8 min. As noted earlier, after each slide *S*s marked a yes or no and a level of confidence on the answer sheet. Also, if an *S* knew the person in a particular slide he so indicated on his answer sheet (a known response was given in less than 1% of the responses in the entire experiment). In order to make sure the test series slides and the numbers on the answer sheets corresponded, *E* called out the number of the slide coming up next when that slide designated either a new column on the answer sheet or the first slide after a tray change.

At the end of the session, *S*s filled in the protocol sheet described above.

Results

The yes and no responses to the target picture in the test series are referred to as hits and misses, respectively. Similarly, the yes and no responses to the 149 decoys are false alarms (FA) and correct rejections (CR).

In order to take into account the confidence levels, two scales were developed and used in evaluating performance. The first scale was based on hits and misses and consisted of 1 for certain miss, 2 for probably miss, 3 for possibly miss, 4 for possibly hit, 5 for probably hit, and 6 for certain hit. The second scale was based on FAs and CRs and consisted of 1 for certain FA, 2 for probably FA, 3 for possibly FA, 4 for possibly CR, 5 for probably CR, and 6 for certain CR. The higher the score, the more accurate the response for either scale. The maximum hit-miss (H-M) score was 6, and the maximum FA-CR score (for 149 decoys) was 894.

The mean H-M score for the eight treatment conditions (collapsed across targets) is presented in Table 1. An analysis of variance indicated a significant main effect for target position ($F = 10.80$, $df = 1, 112$, $p < .005$) with performance better when the target appeared in Position 40. The effect of target exposure duration was slightly less than the usually accepted .05 level ($F = 3.30$, $df = 1, 112$, $.05 < p < .10$) with a higher level of performance at the longer exposure duration. For practical purposes, however, this effect may be worth noting.

TABLE 1

MEAN HIT-MISS SCORES

Observation instructions	Position 40		Position 140	
	10 sec.	32 sec.	10 sec.	32 sec.
Prior knowledge	3.6	4.9	3.0	3.0
No knowledge	3.5	4.4	2.6	3.1

The mean FA-CR scores are presented in Table 2. The effect of observation instructions was significant ($F = 3.92$, $df = 1, 112$, $p < .05$) with better performance in the prior knowledge condition. The Observation Instructions × Target Exposure Duration × Target Position interaction was also significant ($F = 8.10$, $df = 1, 112$, $p < .005$). The data underlying this interaction indicate that prior knowledge improves the FA-CR score only under those conditions of the other variables where the task is relatively easy (32-sec. exposure, Position 40).

Two interactions involving the target variable were also significant: With the H-M score the Target Exposure Duration × Target Position × Target interaction ($F = 5.06$, $df = 1, 112$, $p < .05$), and with the FA-CR score the Observation Instructions × Target Position × Target interaction ($F = 4.72$, $df = 1, 112$, $p < .05$). While it is not surprising that the

TABLE 2

MEAN FALSE-ALARM-CORRECT-REJECTION SCORES

Observation instructions	Position 40		Position 140	
	10 sec.	32 sec.	10 sec.	32 sec.
Prior knowledge	818	871	844	812
No knowledge	823	758	806	834

TABLE 3

HIT RATE: PERCENTAGE OF *S*s WHO RESPONDED "YES" TO TARGET PICTURE

Observation instructions	Position 40		Position 140	
	10 sec.	32 sec.	10 sec.	32 sec.
Prior knowledge	56	81	44	38
No knowledge	56	69	31	44

results of a facial recognition experiment should be a function of the idiosyncratic features of the target, the data underlying these two interactions were not readily interpretable.

While the above analyses take into account the confidence levels of the responses, a more direct measure (for purposes of quick interpretation) is simply the percentage of *S*s who had hits (i.e., said yes when the target picture appeared in the test series). The overall hit rate was 52.2%. The hit rate for the eight conditions of observation instructions, target exposure duration, and target position are shown in Table 3.

The total number of FAs was 1,672 or approximately 9% of the decoys. There was a tendency (but only a tendency, since the effect did not approach significance) for FAs to be correlated with hits (i.e., those *S*s who had hits tended to have more FAs). Only four *S*s had perfect H-M and FA-CR scores.

Although no attempt was made to assign male and female *S*s equally to the various treatment conditions, *S*s of each sex were fairly evenly distributed across the conditions. In order to get a rough idea of possible *S* sex differences, a simple t test was carried out on the H-M scores. The difference just missed being significant at the .05 level, $t = 1.62$, $df = 112$, $.05 < p < .10$. The actual hit rates were 57% and 45% for women and men, respectively.

The protocols were analyzed to determine which facial feature *S*s said they used in the recognition process. A frequency count was made of the number of times various features were mentioned by the different *S*s (i.e., the number of different *S*s who mentioned the feature). The results of the frequency distribution, collapsed across all experimental factors, were general structure, 73; eyes, 66; nose, 53; skin, 42; mouth, 33; lips-chin, 31; hair, 28; and ears, 19.

Discussion

The effects of target exposure time (which was near significance at the .05 level) and target position were as expected. The longer time to study the four candid-position slides of the target probably leads to a better representation in memory that in turn results in better recognition. This finding, incidentally, is consistent with results reported by Leckart (1966) and Leckart, Keeling, and Bakan (1966). This improved memory representation may take several forms such as a stronger memory for particular features or a greater variety of features.

The effect of target position is certainly consistent with most of the psychological literature on memory. The present study does not, of course, separate the two most likely explanations for the effect, namely, decay due to the passage of time and interference due to more intervening items between the initial exposure to the target and the appearance of the target in the sequence.

Although the observation instructions variable, prior knowledge or no knowledge, affected the FA-CR score, it did not significantly influence the hit rate. The lack of an effect on hit rate may simply be due to the fact that the intent of this experimental manipulation was not achieved. The laboratory situation was sufficiently artificial that *S*s in the no prior-knowledge condition gave as much attention to the candid photographs as *S*s in the knowledge condition—regardless of what *E* said to them in the instructions. The effect on the FA-CR score without a corresponding effect on the H-M score (which should be the case if the variable was simply producing a lower criterion for a yes response) is difficult to interpret. The significant Observation Instructions × Target Exposure Duration × Target Position interaction seems to indicate that prior knowledge is useful only if the exposure duration is long enough (32 sec.) to acquire more information about the target and if the target position (Position 40) is early enough to use the information. Again, however, it is not clear why the effect should be reflected in the FA-CR score and not the H-M score.

EXPERIMENT II

The second experiment was designed to explore the effects of pose position on facial recognition performance. Pose position refers to the view of the person in the test-series photographs.

Method

Subjects. The *S*s were 128 students from the same introductory psychology course described in Exp. I. There were 55 males and 73 females.

Design. A 4 × 4 factorial design was employed with both factors manipulated as between-*S* variables. The first factor, pose position, represented the view of the faces in the test series slides. The four levels were front, profile (left side), right portrait, and left portrait. The portrait positions were taken with the camera at an angle of 45° from the front. The second variable, target, simply consisted of four different targets, all 17 or 18 yr. old. There were eight *S*s in each of the 16 treatment conditions.

Procedure. The task (with prior-knowledge instructions), materials, and apparatus were essentially the same as Exp. I. After being told the nature of the task, *S*s were shown four candid pictures of one of the targets for a total time of 32 sec. Answer sheets were handed out and the remaining instructions presented. The *S*s then saw the 150-slide sequence. The slides were drawn from the same library and presented at the same rate described in Exp. I. Of course, in one condition all 150 views were front bust, in another condition all were profile, and in the other two conditions views were left portrait or right portrait. The target's picture appeared only once, in the Position 104. Two of the targets were the same as those used in Exp. I. A protocol was obtained at the end of the test series.

Results

The H-M and FA-CR scores were again calculated. The mean scores are presented in Table 4. Analyses of variance with these two scores revealed no significant effects for pose position, target, or their interaction. Looking at the hit rates, the percentage of *S*s who said yes when the target picture appeared was: front, 53%; profile, 33%; left portrait, 59%; and right portrait, 41%. The overall hit rate was 48%.

TABLE 4

MEAN H-M AND FA-CR SCORES FOR VARIOUS POSE POSITIONS

Pose position	H-M	FA-CR
Front	3.6	844
Profile	2.9	836
Left portrait	3.6	827
Right portrait	2.8	853

Note.—H-M = hit-miss, FA-CR = false alarm-correct rejection.

A *t* test on H-M scores (collapsed across pose position and target) was carried out to compare performance of male and female *S*s. The difference was significant ($t = 2.5$, $df = 126$, $p < .05$) with males performing better than females; means were 3.7 and 2.8 for males and females, respectively. The hit rates were 56% and 42%, respectively. It should be noted that this difference is in the opposite direction of Exp. I, which was not quite significant at the .05 level.

A frequency count was made on the number of times different facial features were mentioned in the protocols. The distribution results were: eyes, 87; nose, 77; lips-chin, 54; general structure, 51; hair, 43; skin, 41; mouth, 25; and ears, 12.

It is also possible to make a comparison between Exp. I and Exp. II with respect to target position in the test series. Taking the front position and appropriate two targets from Exp. II and the 32-sec. exposure time and prior-knowledge conditions from Exp. I, there were three conditions that were alike except for the position of the target picture in the test series. The three were Positions 40, 104, and 140, and the hit rates were 75%, 53%, and 41%, respectively.

Discussion

When this experiment was originally conceived, it was anticipated that the portrait positions might lead to better recognition performance than either the front or profile views. The reasoning was simple; in the portrait pose a combination of features could be observed that was not available in the front or profile poses. It was further speculated that the front pose would be better than the profile for the same reason. Obviously, the data do not support the first hypothesis, although the hit rates do indicate that recognition is poorer with the profile pose (not at the .05 significance level, however). It is possible, of course, to suggest reasons why the different poses do not produce differences in recognition performance. It may be that *S*s use different facial features with the different pose positions, but they may be constrained by some overall limitation on how much information they can handle. One thread of support for this reasoning is that in the protocols different features were mentioned with different frequencies in the various pose conditions. For example, nose was mentioned more frequently in the profile conditions.

EXPERIMENT III

Experiment III was a straightforward study in which color slides were compared to black and white (B & W) slides. It seemed reasonable to speculate that the identification process would be enhanced by the use of color as compared to B & W, because color slides provide a dimension of information nonexistent in B & W.

Method

Subjects. The *S*s were 32 students at the State University of New York at Buffalo. Each *S* was paid $2.00 for participating. There were 17 males and 15 females.

Procedure. With two exceptions, the procedure was the same as Exp. I and II. Instead of looking at four candid-position slides, *S*s exposure to the target consisted of the target person entering the room (live) and handing out the answer sheets. Prior to observing the target, *S*s were instructed as to the nature of the task (i.e., they were told that they would later be attempting to identify the target in a series of pictures). The target stayed in the room for approximately 30 sec. The second difference was that in this study the target's picture appeared twice in the 150-item test series—in Positions 40 and 140. All test series slides were front view, bust length.

The type of photograph variable (color or B & W) was manipulated between-*S*s, while the target position (Position 40 or 140) was within *S*s. There were two targets (one with glasses and one without); this variable was also manipulated between *S*s.

The color and B & W slides were exactly the same. This identity was accomplished when the original library of slides was made by using two cameras mounted side by side (one with color and one with B & W film) and taking the pictures simultaneously.

Results

The H-M and FA-CR scores were computed. An analysis of variance showed no significant effects with the H-M scores. In fact, the mean H-M score for all four conditions of type of photograph and target position was exactly the same, 4.9. A closer look at the target position variable revealed that most *S*s gave the same response when the target occurred the second time (Position 140) as they had given on its first occurrence. Only six *S*s responded differently to the two occurrences of the target, and only one of these differences was more than one point on the H-M scale (it was a difference of 2).

An analysis of the FA-CR scores also showed no significant effects. The means for color and B & W were 872 and 856, respectively.

The overall hit rate was 84%. The hit rates with color and B & W were 87% and 81%, respectively. A comparison can be drawn between the hit rates in this study when the target picture was in Position 40 (84%) and the hit rate in Exp. I in the prior-knowledge condition, Position 40, and the 32-sec. exposure-duration condition (81%).

An analysis was carried out comparing H-M scores for male and female *S*s. The means were males, 4.7 and females, 5.2, a difference that did not approach significance.

As in Exp. I and II, the protocols were analyzed to determine which facial features were most frequently mentioned. The features and their frequencies were: eyes, 24; hair, 18; nose, 11; general structure, 10; skin, 8; and mouth, 8. There were no obvious differences in the feature frequencies between the color and B & W conditions.

Discussion

If color pictures do contain more information than B & W pictures (and they surely must), this difference in information does not lead to better recognition performance.[3] It seems reasonable to speculate that the failure to find a difference, as with pose position in Exp. II, can be attributed to the information processing limitations of *S*s. (They can only deal with a certain amount of information in a given period of time.)

The almost identical responses to the two occurrences of the target picture is interesting in that once *S* identifies the target picture, he does not make a mistake later. On the other hand, if *S* misses the target on the first occurrence, he is not likely to recognize it later.

---

[3] The color versus B & W slide variable has been manipulated as part of another experiment and replicated the no-difference finding of this study.

GENERAL DISCUSSION

It is interesting to note that even in the most ideal conditions of the present experiments, the hit rates (81% and 84% in Exp. I and III, respectively) were lower than the 90% and higher rates previously reported. Granted, the task was different from some of the earlier studies where *S*s were choosing between a pair of pictures instead of making an absolute judgment on individual pictures. Although, in the Nickerson (1965) experiment, a similar absolute judgment was made with higher performance results. However, Nickerson used pictures of "things," which indicates that unfamiliar faces may be more difficult to recognize than other types of visual stimuli. Taking into account these procedural differences the earlier studies may be less relevant to the criminal identification problem than previously thought.

Two of the findings in the present experiments do seem reasonably clear. The more pictures an *S* (or witness) looks at before seeing the target picture, the less likely he is to recognize the target. Also, the better the opportunity to study the target, the greater the probability of a correct identification. While the significant interactions involving these variables indicate some qualifications regarding their effects, the findings are important. If law enforcement agencies can find ways of organizing files so that the witness has fewer alternatives to search, the result could be improved identification rates.

The fact that neither pose position nor type of photograph (color or B & W) affected performance was somewhat unexpected, since it was originally thought that the additional information in color slides and portrait poses would enhance recognition performance. As already mentioned, the negative findings with these variables may be the result of *S*s' inability to make use of additional information. In any case, the outcome suggests that the present B & W, front and profile views used by law enforcement agencies need not be changed to color or a portrait pose (at least not for reasons of improving identification rates).

The conflicting results of *S* sex in these experiments is currently a mystery. Perhaps further research will define which sex, if either, makes the better witness.

The facial features mentioned in the protocols were reasonably consistent across the three experiments. Although these feature data are interesting, they should only be considered tentative or exploratory. The question of which features *S*s use in the identification process is an important issue, however, and deserves more extensive and elaborate investigation.

Finally, a comment on the generality of the results seems in order. The null finding with regard to target differences in Exp. II would seem to argue that the results do apply to some portion of the population. Most likely, however, the finding depends on the fact that the targets and decoys represent a sampling from the middle range with respect to general appearance and specific physiognomic traits. Surely, *S*s can be expected to recognize any highly distinctive face. However, in the actual identification of criminals it is the nondistinctive face that presents the problem, and these studies were planned accordingly.

REFERENCES

BOWER, G. H. Mental imagery and associative learning. In L. W. Gregg (Ed.), *Cognition in learning and memory*. New York: Wiley, 1970.

HOCHBERG, J. & GALPER, R. E. Recognition of faces: 1. An exploratory study. *Psychonomic Science*, 1967, **9**, 619–620.

LECKART, B. T. Looking time: The effects of stimulus complexity and familiarity. *Perception and Psychophysics*, 1966, **1**, 142–144.

LECKART, B. T., KEELING, K. R., & BAKAN, P. The effect of rate of presentation on free looking time. *Perception and Psychophysics*, 1966, **1**, 107–109.

NICKERSON, R. S. Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology*, 1965, **19**, 155–160.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

SHEPARD, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 156–163.

YIN, R. K. Looking at upside-down faces. *Journal of Experimental Psychology*, 1969, **81**, 141–145.

(Received May 7, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 484–486

CHANGES IN AUTHORSHIP, AFFILIATION, AND RESEARCH SUPPORT PATTERNS OF *JOURNAL OF APPLIED PSYCHOLOGY* ARTICLES, 1917–1969

GARY N. CHAISON[1]

State University of New York at Buffalo

An analysis of articles from the *Journal of Applied Psychology* was made in regard to variables of multiple authorship, funding sources cited by authors, and author affiliations. The sample of 729 articles consisted of every fifth article appearing in the Journal from 1917 to 1969. Results indicate trends toward increased multiple authorship, more extensive and diverse funding, and changes in the nature of author affiliations.

The scientific journal as a primary method of communication between researchers may provide a source of information on identifiable trends in the research process as well as the changing attributes of researchers. The purpose of this article is to examine some author and funding characteristics over a long time span for a particular journal, the *Journal of Applied Psychology*. Specifically, changes in authorship and coauthorship patterns, sources of funding for the research reported, and institutional affiliation of the researchers were examined.

The study of psychological journals seems to have received considerable research attention. Walsh, Feeney, and Resnick (1969) investigated the graduate school origins of authors of articles from a psychological journal, while Smith (1958) and Mitchell (1961) focused upon the increasing trend toward multiple authorship. Xhignesse and Osgood (1967) studied the use of biographical citation as an indication of the nature of the communication network among psychological researchers. Woods (1961) studied the APA divisional membership of article authors, while author institutional affiliations were examined by Goodstein (1963), Golightly, Byrne, and Capaldi (1963), and Bohn (1966).

The preceding studies have been strongly empirical in nature, basically extracting data from sampled articles or volumes for the selected variables. In contrast to this orientation, Darley (1968) examined the *Journal of Applied Psychology* and presented an historical overview of the development of the Journal as well as the field of applied psychology. Writing his article to note the fiftieth anniversary of the founding of the Journal, Darley read every fifth volume in its entirety, recording the early emergence of seemingly contemporary methodology and the often well-deserved fall of certain tests into obscurity. In his conclusion he suggests that applied psychologists have a "cavalier attitude" toward their own history (Darley, 1968, p. 9). The purpose of the present article is to provide some additional historical perspectives on the sources of authorship and research reported in the Journal.

METHOD

The study uses a sample of every fifth article, selected after a random start, from the first 53 volumes of the *Journal of Applied Psychology*. This covers Volumes 1–53 for the years 1917–1969. Essays, comments, and letters have been excluded from the sample.

Designating time as the independent variable, data were tabulated for the three following dependent variables: (*a*) number of authors per article; (*b*) source of funding, if any, cited by the author(s) of the articles; (*c*) type of institutional affiliation cited by the authors of the articles.

A total of 729 articles were sampled, including the works of 824 individual authors. The data were compiled by year but then combined into periods of 5 yr. (after an initial time period of 3 yr. for 1917–1919). This was done because of the small number of articles sampled during some years (often less than 10 articles per year).

---

[1] Requests for reprints should be sent to Gary N. Chaison, Department of Industrial Relations, State University of New York at Buffalo, Crosby 319A, Buffalo, New York 14214.

RESULTS

Number of Authors per Article

Figure 1 indicates a trend toward collaboration in applied psychological research. During the latest time period almost half of the sampled articles were written by two or more persons.[2] The causes of this noticeable trend may be numerous. Smith (1958) has suggested that factors affecting collaboration in published research might be: (*a*) the type of problem entailed in the research, that is, applied research usually involves group efforts as opposed to more individualistic theoretical works; (*b*) method of data collection and analysis, that is, at times the burdensomeness of the data-collection task or the method of analysis applied by the researcher may require collaboration; (*c*) affiliations of authors, that is, authors in large institutions may be more likely to collaborate because of the nearness of colleagues with similar research interests; and (*d*) support by government funds or foundations, that is, the use of project or team organization—often necessary to utilize substantial funding—may affect the extent of collaboration.

Furthermore, Mitchell (1961) has suggested that there may be a trend, reinforced by psychology's code of ethics, toward giving students and assistants credit for their work in published form.

Types of Funding Cited by Authors

The emergence of funding for research from both governmental and private sources may be seen in Figure 2. The percentage of articles in which there is no citation of funding source has rapidly and steadily decreased since the early 1940s. However, the outstanding feature of this figure is the large proportion of research in the Journal for which no financial support is cited. Despite the large outlay of government and private support, most of the published work in the Journal does not receive outside support. The curves show the increase over the years in both DOD and HEW agency support and reflect the substantial fluctuations in government funding policies.

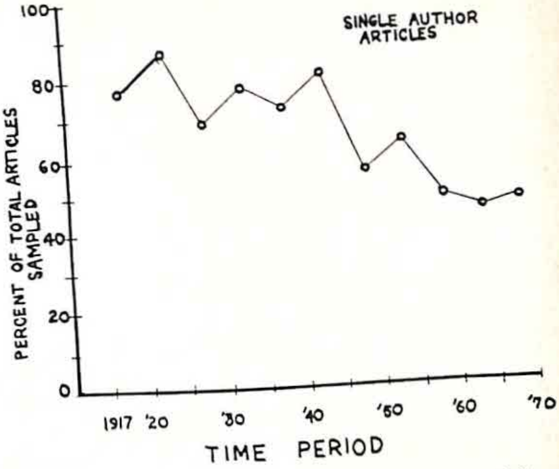


FIG. 1. Multiple authorship of sampled articles.

Author Institutional Affiliation

Figure 3 indicates the types of institutional affiliations of the authors of the articles. In the sample of 729 articles, 1,046 author affiliations were recorded because of multiple

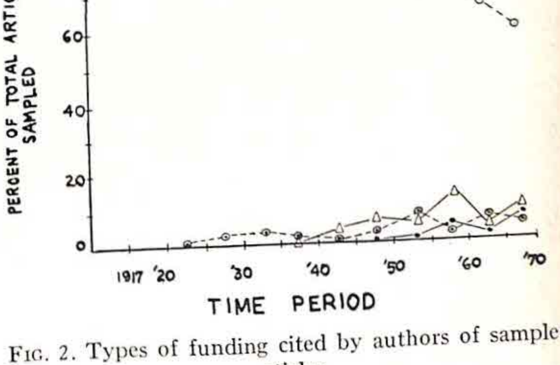


FIG. 2. Types of funding cited by authors of sampled articles.

---

[2] More complete, tabular breakdowns on which this and subsequent figures are based may be obtained from the author. (See address in Footnote 1.)

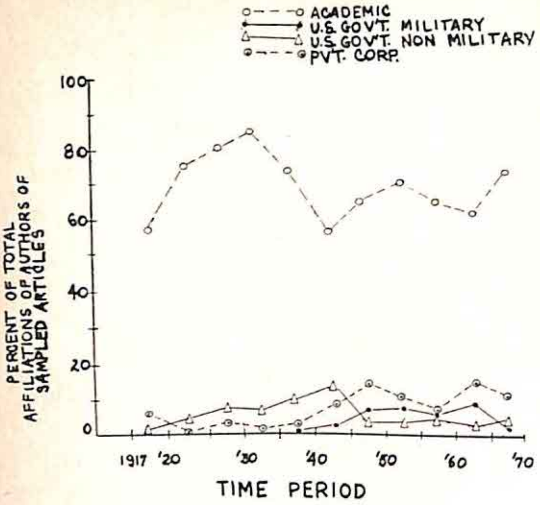


FIG. 3. Institutional affiliations of authors of sampled articles.

authorship. However, because some author's works appeared more than once in the sample, the number of individual authors was actually 864. In the tabulation, the affiliations of all authors of sampled articles were recorded, hence, the total of 1,046 affiliations.

As Figure 3 illustrates, the proportion of articles from academic institutions exceeded 80% in the early 30s. The percentage in the academic category has remained relatively stable at around 60%–70% since 1940. Other types of affiliations cited have fluctuated more with some trend of increasing contributions from industrial affiliations. This source reached a high point during the 1960–1964 period when more than one-fifth of the authors cited industrial affiliations. This category includes industry as well as private research institutions, both nonprofit and profit. Figure 3 also indicates that governmental nonmilitary affiliations of authors have declined since the peak in the early 1940s, while there is an increase in authors citing military affiliations since those years, at least until the mid 60s. It should be pointed out that much of the government-supported research reflected in Figure 2 is carried out under grants and contracts with individuals affiliated with private research and academic institutions. Hence, the differences between Figures 2 and 3.

The *Journal of Applied Psychology* has had only six different editorial administrations in its long history. Changes in Editor occurred in 1917, 1921, 1943, 1955, 1961, and 1971. It would be difficult to relate these trends and fluctuations to changes in editorial policies.

CONCLUSION

The preceding descriptive analysis of articles in the *Journal of Applied Psychology* from 1917 to 1969 indicates the following trends: (*a*) a greater tendency toward collaboration in applied psychological research, (*b*) increased diversity in the sources of funding cited by article authors. There still appears to be, however, a small amount of funded research relative to the amount of unsupported research reported. (*c*) An increased diversity in types of author affiliations as well as a greater role for private organizations as a source of applied psychological research were also apparent. A large majority of authors, however, still come from academic institutions.

REFERENCES

BOHN, M. J., JR. Institutional sources of articles in the *Journal of Counseling Psychology*—Four years later. *Journal of Counseling Psychology*, 1966, **13**, 489–490.

CRANE, D. The gatekeepers of science; Some factors affecting the selection of articles for scientific journals. *American Sociologist*, 1967, **2**, 195–201.

DARLEY, J. G. 1917: A journal is born. *Journal of Applied Psychology*, 1968, **52**, 1–9.

GOLIGHTLY, G., BYRNE, D., & CAPALDI, E. J. Publication of psychological research among PhD-granting institutions: 1952–1961. *Perceptual and Motor Skills*, 1965, **20**, 87–95.

GOODSTEIN, L. D. The institutional sources of articles in the *Journal of Counseling Psychology*. *Journal of Counseling Psychology*, 1963, **10**, 94–95.

MITCHELL, M. B. Trends toward multiple authorship in scientific publications. *Journal of Psychology*, 1961, **52**, 125–131.

SMITH, M. The trend toward multiple authorship in psychology. *American Psychologist*, 1958, **13**, 596–599.

WALSH, W. B., FEENEY, D., & RESNICK, H. Graduate school origins of *Journal of Counseling Psychology* authors. *Journal of Counseling Psychology*, 1969, **16**, 375–376.

WOODS, P. J. Some characteristics of journals and authors. *American Psychologist*, 1961, **16**, 699–701.

XHIGNESSE, L. V., & OSGOOD, C. E. Bibliographical citation characteristics of the psychological journal network in 1950 and 1960. *American Psychologist*, 1967, **22**, 778–791.

(Received June 8, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 487–488

SHORT NOTES

THE VALIDITY OF IMPORTANCE[1]

MILTON R. BLOOD[2]

University of California at Berkeley

Previous investigators have questioned the usefulness of weighting by importance when combining measures of satisfaction with job aspects into a measure of overall satisfaction. Data are presented that call into question the validity of the information that is used for importance weighting. The model of cognitive behavior that relates the importance of job aspects to the evaluation of overall job attitude should be revised.

The importance of various job aspects to workers has been considered by some investigators to be an influential variable in the determination of overall job satisfaction (Friedlander, 1965; Glennon, Owens, Smith, & Albright, 1960; Weiss, Dawis, England, & Lofquist, 1964a, 1964b). On the other hand, others have shown that weighting satisfaction components by their importance ratings did not increase the fit of component measurements to measurements of overall job satisfaction (Ewen, 1967; Schafer, 1953) or to turnover (Mikes & Hulin, 1968).

The general model that has been used to relate importance measures to satisfaction measures is $S_o = \Sigma\, I_a S_a$, where S_o = overall job satisfaction, I_a = importance of a job aspect, and S_a = satisfaction with a job aspect. Such a model allows predictions of overall satisfaction to be based on the combination of component satisfactions with the components being idiosyncratically weighted by individual workers. The general advantage of using such a model has been discussed by Vroom (1964, pp. 163–165).

If the information given by workers in ratings of the importance of various job aspects is the same as that used by the workers in producing a rating of overall satisfaction, then that information should be discernible in the relationships between satisfaction with the job aspects and overall job satisfaction. The aspect a worker says is most important to him should be the aspect most highly related to his overall job satisfaction; that aspect ranked second in importance should be second most related, and so on to the aspect with the least importance, which should have the smallest relationship with his overall job satisfaction.

Method. The *S*s for this study were 380 clerical workers from the home offices of a large corporation. Among the measures that *S*s completed as part of a larger study were the Job Descriptive Index (JDI; Smith, Kendall, & Hulin, 1969), which measures satisfaction with five aspects of the job (work itself, coworkers, supervision, pay, and promotions); a measure of overall job satisfaction on a Faces scale (Kunin, 1955); and a ranking of the importance of the five job aspects.

For the entire group, the satisfaction scores on each of the five JDI job aspects were standardized. Then, for each *S*, the standardized aspect scores were reordered so that the first variable for each individual was the standardized satisfaction score on the aspect that he ranked as most important; the second variable was the standardized satisfaction score on the aspect that he ranked second most important; and so on to the last of the aspect variables that was the standardized satisfaction score on the aspect ranked fifth in importance. The satisfaction variables from the JDI were thus no longer designated by content areas. Instead, the variables were defined by the idiosyncratically assigned importance levels. For each *S*, the first variable was his standardized satisfaction score for the job aspect that he ranked as most important; his second variable was the standardized satisfaction score for the job aspect that he ranked second in importance, etc., through all five job aspects.

Very little distortion in scale units resulted from the reordering. Since the original scales were standardized before the reordering, they all had a mean of 0 and a standard deviation of 1. The means of the reordered scales ranged from −.04 to +.04, and their standard deviations ranged from .98 to 1.03.[3]

In determining the relative influence of multiple predictors, Darlington (1968) suggests consideration of three different sets of information: (*a*) the multiple-correlation beta weights; (*b*) the zero-order correlation coefficients; and (*c*) the "usefulness" of the predictors, that is "the amount that the squared multiple correlation would drop if the variable were removed . . . (p. 180)." All three indexes (of the relative influence of the aspect satisfaction measures) were computed with the reordered measures of satisfaction with job aspects as predictors and the measure of overall job satisfaction as the criterion.

---

[1] The author wishes to thank Charles L. Hulin and Patricia Smith Mikes for making available the data that were analyzed for this study.

[2] Requests for reprints should be sent to Milton R. Blood, Department of Psychology, University of California at Berkeley, Berkeley, California 94705.

[3] Equivalence of variance among the scales is an important characteristic since a lack of variance would reduce the correlation involving even a variable with high causal relevance.

Results. The multiple-correlation coefficients with the job aspect satisfaction variables (as reordered for importance) as predictors and overall job satisfaction as the criterion was .43 ($p < .001$). The beta weights of the measures of satisfaction with job aspects in descending order from most important to least important were .16, .10, .16, .03, and .18. The zero-order correlations of the job aspects with overall job satisfaction in descending order from most important to least important were .31, .27, .30, .21, and .32. The measure of "usefulness" suggested by Darlington (1968) is proportional to the squared partial correlations between each predictor and the criterion, holding all other predictors constant. The squared partial correlation coefficients for the aspect variables in descending order of importance were .02, .01, .03, .00, and .03. For the sake of comparison, it is interesting to note that predicting overall satisfaction in this sample from the job aspect variables defined in the usual manner (by their content names) gave a multiple-correlation coefficient of .45 ($p <$ .001), almost exactly the same as the reordered variables.

Discussion. If the importance rankings given by *S*s are related to the actual importance of the aspects as they enter into *S*s' considerations in rating overall satisfaction, then the validity coefficients, beta weights, and usefulness indexes for the aspect variables all should have been in descending order from most important (first variable) to least important (fifth variable). None of the indexes achieved the appropriate order.

In Schafer's (1953) sample, overall satisfaction had a significantly higher correlation with the aspect ranked idiosyncratically most important than with the aspect ranked least important (twelfth). The relationship was not regular, however, for Variables 1–12. The rank-order correlation coefficient between importance and correlation with overall satisfaction was not significant. The variables ranked first, second, third, ninth, and eleventh in importance were significantly related to overall satisfaction, while the variables ranked sixth and seventh had the lowest relationships. Ewen (1967) found that the correlation between overall satisfaction and satisfaction with the most important aspect was significantly higher than the correlation of overall satisfaction and satisfaction with the least important aspect in only one of three samples, when using the five job aspects measured by the JDI. (In all three samples the difference between the correlations was in the predicted direction.)

Though satisfaction with an aspect of a job may vary in strength of relationship with overall job satisfaction, this strength of relationship has little, if any, relationship to the importance rank that an individual will assign to that aspect. Of course, there must be some aspects that are totally irrelevant. Previous investigators of the importance of job aspects seem to have been sensible men and have excluded such dimensions, but the question of relevance can be answered empirically in the case of importance. We should consider any aspect of the job as unimportant in the determination of overall satisfaction if satisfaction with that aspect does not correlate at some predetermined level with overall satisfaction. Such variables should be given a weight of zero (in effect dropping them from the sum of aspects considered). Use of measurements of importance for differential weighting is not warranted since these measures are not related to the strength of the relationship between the aspects and overall satisfaction. A possible manner of operation is to treat importance as a binary property with all "important" aspects getting equal weight. The general formula, $S_o = \Sigma\ I_a S_a$, would be retained, but the variable I_a could take on only the value of 0 or 1.

Though this is the same conclusion to which Ewen (1967) came (to use unit weights rather than weighting by importance), the present argument is based on an additional premise. Ewen argued that differential weights did not make a difference (or, if you prefer, that "importance" is not important). The present argument adds the consideration that when making judgments about the importance of job aspects, *S*s provide information that has little relationship to the actual influence of those aspects on an overall evaluation of the job. It is unwarranted to expect people to behave in the methodical and orderly manner necessary to provide data that would support the model of behavior implied by the original properties of the formula. That conception of importance is empirically invalid.

REFERENCES

DARLINGTON, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin*, 1968, **69**, 161–182.

EWEN, R. B. Weighting components of job satisfaction. *Journal of Applied Psychology*, 1967, **51**, 68–73.

FRIEDLANDER, F. Relations between the importance and the satisfaction of various environmental factors. *Journal of Applied Psychology*, 1965, **49**, 160–164.

GLENNON, J. R., OWENS, W. A., SMITH, W. J., & ALBRIGHT, L. E. New dimensions in measuring morale. *Harvard Business Review*, 1960, **38**, 106–107.

KUNIN, T. The construction of a new type of attitude measure. *Personnel Psychology*, 1955, **8**, 65–78.

MIKES, P. S., & HULIN, C. L. Use of importance as a weighting component of job satisfaction. *Journal of Applied Psychology*, 1968, **52**, 394–398.

SCHAFFER, R. H. Job satisfaction as related to need satisfaction in work. *Psychological Monographs*, 1953, **67**(14, Whole no. 364).

SMITH, P. C., KENDALL, L. M., & HULIN, C. L. *The measurement of satisfaction in work and retirement*. Chicago: Rand-McNally, 1969.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

WEISS, D. J., DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. Construct validation studies of the Minnesota Importance Questionnaire. *Minnesota Studies of Vocational Rehabilitation*, 1964, **16**, Bulletin 41. (a)

WEISS, D. J., DAWIS, R. V., ENGLAND, G. W., & LOFQUIST, L. H. The measurement of vocational needs. *Minnesota Studies of Vocational Rehabilitation*, 1964, **16**, Bulletin 39. (b)

(Received May 22, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 489–490

RELATIONSHIP OF INITIATING STRUCTURE AND JOB PERFORMANCE AS MODERATED BY CONSIDERATION

ROBERT C. CUMMINS[1]

Department of Business Administration, University of Texas at Arlington

Many researchers have suggested that additional variables are necessary to understand the relationship of leader behavior and group performance. The present study investigates the hypothesis that the effect of initiating structure on the rated productivity and quality of work-group performance in an industrial manufacturing plant is moderated by considerate leader behavior (rated and self-report). The results support the hypothesis when quality is the criterion measure but do not support it when productivity is the criterion measure. The results are interpreted as being consistent with Fleishman and Harris' finding that grievances and turnover are less affected by initiation of structure when foremen are rated high on consideration.

Consideration and initiating structure, two of the dimensions of leadership behavior identified by the Ohio State University research programs on leadership (cf. Halpin & Winer, 1957), have received considerable attention in the literature. Despite the amount of research on these two variables, Korman (1966) concludes a literature review by stating that "very little is known as to how these variables may predict work group performance and the conditions which affect such predictions [p. 360]." Fleishman and Harris (1962) suggest that the failure to consider the interaction effects of consideration and initiating structure may account for the inconsistent research findings. The results of their study indicate that consideration moderates the effect of initiating structure on grievances and turnover. More specifically, their data suggest that high-consideration foremen can increase structure with very little increase in grievances and no increase in turnover, whereas, structure is directly related to these indexes for low-consideration foremen.

The purpose of the present note is to determine whether consideration moderates the effect of initiating structure when productivity and quality are the criterion measures. It is predicted that initiating structure will be more closely related to measures of productivity and quality for high-consideration foremen as compared to low-consideration foremen. This hypothesis is based on the Fleishman and Harris (1962, pp. 53–54) reasoning that high-consideration foremen may establish a climate of mutual trust and that in such a climate, workers are more likely to accept (and implement) structure initiated by the leader.

Method. The *S*s were 133 foremen at a large manufacturing plant in the midwest employing about 3,500 nonsalaried workers. The average work group size was 21 and supervised activities ranged from unskilled to highly skilled engineering and machining tasks.

The *S*s were administered the Leadership Opinion Questionnaire (LOQ) (Fleishman, 1969) as part of a large battery of tests aimed at providing information for manpower planning. The LOQ measures self-reported supervisory attitudes scored on consideration and structure. They were informed that this test information would not be available to their immediate superiors.

Ratings of work-group productivity (overall productivity of the work group), quality (amount of scrap, rework, and errors kept to a minimum), leader's initiation of structure (supervisor directs the group, plans and explains things in detail, figures out what to do in new situations), and consideration shown by the leader (supervisor tries to get along with his men, talks with them, shows respect for their ideas and feelings) were obtained.

Each foreman was rated on 7-point Likert scales by at least four managers who were above the ratee in the organizational hierarchy and had worked with the ratee for a long period of time, usually several years. The ratings were administered to each rater separately in a face-to-face interview. This administration allowed the researcher to check each rater's understanding of the dimensions and purpose of ratings. On the assumption that halo effect would be minimized, managers were instructed to rate all foremen on one scale before rating any foremen on the second scale.

Results. The interrater reliability of the ratings was estimated by use of the Spearman-Brown prophecy formula correction of the split-half correlation. The split-half correlation was obtained by dividing the ratings concerning a particular foreman into two sets of ratings and averaging; these two averages were obtained for all foremen rated, and the correlation between the averages was computed for each scale. The interrater reliability coefficients for the four rated dimensions ranged from .59 to .63.

The correlation between ratings of leader behavior and self-reported supervisory attitudes of the same dimensions were low; therefore, ratings and self-

---

[1] Requests for reprints should be sent to Robert C. Cummins, Department of Business Administration, University of Texas at Arlington, Arlington, Texas 76010.

TABLE 1

CORRELATIONS BETWEEN STRUCTURE AND RATED PRODUCTIVITY AND QUALITY, AS MODERATED BY CONSIDERATION

Correlation	High consideration (Rated)[a]	Low consideration (Rated)[b]	High consideration (LOQ)[b]	Low consideration (LOQ)[a]
Structure (Rating) r Productivity	.62	.59	.68	.61
Structure (LOQ) r Productivity	.04	−.26	−.20	−.03
Structure (Rating) r Quality	.59**	.21**	.53*	.18*
Structure (LOQ) r Quality	−.07†	−.36†	−.25	−.32

Note.—LOQ = Leadership Opinion Questionnaire.
[a] $N = 62$.
[b] $N = 71$.
* These two correlations differ at the .05 level of confidence.
** These two correlations differ at the .01 level of confidence.
† These two correlations differ at the .01 level of confidence.

reports on the same behavioral dimensions are treated as two separate variables in the analysis.

As can be noted in Table 1, the results support the hypothesis that rated initiation of structure would be more closely related to quality of performance for high-considerate foremen than for low-consideration foremen; this is true where consideration is measured by either the LOQ or by management ratings of behavior. With respect to behavior ratings, high correlations showed that supervisors who also had higher structure had higher quality. There is also some indication in Table 1 that supervisors with attitudes low in consideration had quality go down when more structure was exerted. Neither measure of consideration moderates the relationship between structure and productivity.

This pattern of results may be explained by examining the criterion measures. In most nonsalaried jobs in the surveyed plant, productivity is likely to be influenced by conditions that are not directly related to leader-member interaction. For example, the production foreman's relation with maintenance may influence the speed and/or quality of maintenance on critical machines and thus affect productivity, whereas, in the opinion of many supervisors, quality of performance depends more on worker motivation and, therefore, may be influenced by considerate leader-member interaction.

If this interpretation is correct, then the finding that the Consideration × Structure interaction affects quality but not productivity appears to be consistent with the Fleishman and Harris (1962) finding that this interaction affects grievances and turnover. In this situation, quality, grievances, and turnover may all be related to the workers' evaluation of his situation, whereas, achieving an acceptable level of productivity may be less dependent on positive attitudes toward work.

It is interesting to note that despite the small negative correlation between questionnaire results and ratings of consideration, both measurements act as moderators for initiation of structure. Thus, both considerate attitudes and considerate behavior make other leader behaviors more effective.

A methodological note is in order. Although much effort was expended to reduce the possibility of halo, the ratings may be correlated partly because of method specificity rather than because of some underlying causal relationship. The present analysis that compares correlations at two levels of a third variable should not be influenced by halo. That is, there is no a priori reason to believe that halo has any effect on the differences between correlations.

Conclusion. The structuring behavior of first-line supervisors has a greater effect on rated quality of performance when the supervisor believes he should show consideration and/or is rated high on consideration. Neither attitudes endorsing consideration nor ratings of consideration are moderators of the relation between initiating structure and productivity. These conclusions seem consistent with results of Fleishman and Harris (1962).

REFERENCES

FLEISHMAN, E. A. *Manual for the Leadership Opinion Questionnaire.* (Rev. ed.) Chicago, Ill.: Science Research Associates, 1969.

FLEISHMAN, E. A., & HARRIS, E. Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology*, 1962, **15**, 43–46.

HALPIN, A. W., & WINER, B. A factorial study of the leader behavior description. In R. Stogdill & A. Coons (Eds.), *Leader behavior: Its description and measurement.* Columbus, Ohio: Bureau of Business Research, 1957.

KORMAN, A. K. "Consideration," "initiating structure," and organizational criteria. *Personnel Psychology*, 1966, **19**, 349–361.

(Received October 12, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 491–492

THE APPLICATION BLANK:

HOW HONEST ARE THE RESPONSES?[1]

IRWIN L. GOLDSTEIN[2]

University of Maryland

An examination of the application blank responses provided by a perspective applicant and by the applicant's previous employer indicated that there was substantial disagreement between the two sources. In two categories, duration of previous employment and previous salary earned, the two sources disagreed on 57% of the cases. In these categories, the typical applicant overestimated his duration and his previous salary. Also, there was substantial disagreement on the reason for leaving the previous job. The data appear to support the hypothesis that inaccurate information is likely to appear on the application blank as well as in the interview situation.

Many recent studies (Fleishman & Berniger, 1960; Scott & Johnson, 1967) have demonstrated that the application blank can be used to successfully predict tenure on the job. Scott and Johnson have especially noted the feasibility of employing the application blank in low-skilled occupations. Recently, a firm that was cooperating with the investigator on a research project questioned the honesty of the responses given by the applicants applying for jobs as nurses aides. The honesty of the replies is an issue aside from the question of validity since it is possible to predict from most any set of data. In any case, a search of the literature did not disclose any studies that had collected data reflecting on the honesty of the application blank responses.

Several studies, however, have questioned the accuracy of data collected in the interview. A study by Keating, Paterson, and Stone (1950) found that the correlations between reports on wages, and duration of employment were above .90. Another study by Weiss and Dawis (1960) discovered poorer agreement for a sample of 91 physically handicapped persons. The information varied in accuracy from 0% invalid information for reports on sex to 55% on the questions of receiving assistance. Essentially, invalid information was found for substantial numbers of persons (over 20% of the sample) on questions related to previous education, age at disablement, received assistance, job title, pay, and length of employment.

Blum and Naylor (1968) have suggested that the interview results in a face-to-face pressure that might cause a respondent to stretch the truth. In addition, prospective employees might be more likely to respond honestly on the application blank when their responses are actually recorded in their own handwriting. The present study examined the data recorded on the application blank for a sample of 111 applicants.

Method. The original sample consisted of 173 persons with the final sample equalling 111 persons for whom matching application blank data and previous employer responses were available. The application blanks were filled out by persons applying for positions as nurses aides in a group of nursing homes owned by the same corporation in a large east coast city. The blanks were collected during the period from September 1, 1968, to June 1, 1969. A form was filled out for each of these individuals with the following items listed: name, address, dates employed by last employer, and position previously held. The information for these items was obtained from the completed application blank. This form was sent to the employer listed by the applicant as the last place they were employed. The following information was requested from the previous employer: marital status, position held, dates of employment, salary of employee, and why the employee left. The previous employer was requested to draw a line through the item if he did not have the requested information. He was also asked to return the questionnaire and indicate "not employed by us" if he had never employed the applicant. The previous employers were assured of the confidentiality of the individual pieces of information as well as their firms' anonymity. They were also assured that their information would not affect the employment of the individuals concerned.

Results and discussion. The data generally indicate a substantial number of discrepancies between the employee's application blank responses and the responses of the previous employer. The interpretation of these data do require a note of caution. In this study, two sets of responses are being compared, and the true state of the world is not available. It is possible that some of the previous employer's responses are in error. As the discussion proceeds, however, it will become apparent from some of the directions of the differences that the most parsimonious explanation seems to point to a bias on the part of the applicant.

---

[1] Arnold Richman's aid in collecting the data from the application blanks is gratefully acknowledged. The author also wishes to thank William Mobley and Sharon Dorfman for their analyses and comments about the data.

[2] Requests for reprints should be sent to Irwin L. Goldstein, Department of Psychology, University of Maryland, College Park, Maryland 20742.

For the 111 employers who responded to the survey, 15% (17 respondents) indicated that the applicant had never worked for them. Married females who may have been previously employed under their maiden name were excluded from this count. Those applicants ($N = 94$) whom the employer indicated had worked for their firm were included in the rest of the analyses. Information that was left blank by the applicant or the employer (e.g., in some cases salary disclosure was against company policy) reduced the N for the various categories. Also, in some cases, the respondent used completely different scales (i.e., pay per week versus pay per hour without the number of hours per week being stated). In these cases, or in any other case in which the responses could not be reconciled, the data were not used in these analyses. This was done in order to avoid, wherever possible, reported differences where none might exist.

The largest amount of agreement between the applicant and the previous employer occurred on the topic of the previous position held. There were only five clear discrepancies. Considering that the applicant was applying for a position as an aide in a nursing home, which is an unskilled low socioeconomic-status job, the applicant would be hard pressed to indicate that he or she had previously held a job substantially different from the job being considered. The category with the next highest percentage disagreement was the reason(s) for leaving. In the 63 cases in which the data were comparable, 25% were in disagreement. These are cases in which there were clear-cut differences. For example, the previous employer stated that the applicant was dismissed because of a violation of the company's rules, while the applicant stated that he left because the pay was too low.

The categories in which the discrepancies were largest were the duration of previous employment and the previous salary earned. As far as duration of employment, there was 57% disagreement. Since this category (as well as the salary data to be discussed next) provided information that permitted an assessment of the degree of overestimation or underestimation, the normal approximation to the binomial was used to test the significance. In the duration of employment data, the applicant's response, as compared with the previous employer, overestimated the amount of time spent on the job in 39 of the 46 cases ($z = 4.57$, $p < .01$, two-tailed test). For the six cases in which the applicant underestimated the time spent on the job, the average underestimation equalled 3 mo. per person. Most of that total resulted from one person who underestimated the time he spent on the job by one year. For those who overestimated, the average overestimation was 16 months. An inspection of the data indicate that the average overestimation was not raised substantially by a few applicants but, rather, occurred across most of the applicants.

It was difficult to analyze the degree of estimation data for salaries because the types of differences were more complicated. In the 43 cases in which there was disagreement, there were 18 instances where the applicant and previous employer disagreed on the mode of payment. Most often, the previous employer indicated that the applicant worked on a piece-rate basis while the employees indicated they worked on a dollar-per-hour basis. For the 25 instances in which the data could be compared, the applicant overstated his salary in 18 instances ($z = 4.56$, $p < .05$, two-tailed test).

In summary, there were a substantial number of discrepancies between the applicant's responses and those answers provided by the previous employers. It is interesting to note that this was not true of all applicants. Of the 94 employees that were reported by the previous employer as having been employed, 31 sets of data were in complete agreement. Thus, the remaining 63 employees disagreed with the previous employers on two of the four categories (previous position held, duration of employment, previous salary earned, reason for leaving) that were examined.

These data suggest that inaccurate information is likely to appear on the application blank as well as in the interview. The same face-to-face pressure that Blum and Naylor (1968) discuss in their analysis of the interview situation probably does not exist in the context of the application blank. The pressure to obtain a job, however, may be equally demanding in both situations. The interview data indicate that two of the responses with the most inaccurate information are duration of previous employment and previous salary earned. The data in this note on application blanks indicate that the largest discrepancies also occurred in these two categories.

REFERENCES

Blum, M. L., & Naylor, J. C. *Industrial psychology.* New York: Harper & Row, 1968.

Fleishman, E. A., & Berniger, J. One way to reduce office turnover. *Personnel,* 1960, **37**, 63–69.

Keating, E., Paterson, D. G., & Stone, C. H. Validity of work histories obtained by interview. *Journal of Applied Psychology,* 1950, **34**, 6–11.

Scott, R. D., & Johnson, R. W. Use of the weighted application blank in selecting unskilled employees. *Journal of Applied Psychology,* 1967, **51**, 393–395.

Weiss, D. J., & Dawis, R. V. An objective validation of factual interview data. *Journal of Applied Psychology,* 1960, **44**, 381–385.

(Received July 16, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 493–494

FACTORS RELATED TO ACCEPTANCE AND REFUSAL OF EMPLOYMENT AMONG TEMPORARY HELP APPLICANTS[1]

MARTIN J. GANNON[2] AND URI BRAININ

Behavioral Science Division, Department of Business, University of Maryland

Although the temporary help service industry has been described historically, there are no published studies that have tested specific hypotheses relative to the type of individual who is attracted to this kind of work. The present study analyzed two types of applicants, 299 in all, who applied to a temporary help firm during 1968. Although both types of applicants were interviewed and were hired, one group refused while the other accepted employment. Of 18 personal characteristics, 5 were predictive of acceptance, namely, type of job, number of skills, availability for work in terms of days per week, expressed interest in a permanent job as a temporary worker, and past experience in terms of months. The analysis delineates several disadvantages in the use of temporary workers, especially when the labor market tightens.

Temporary help firms, which hire individuals and then send them out to do temporary work for any company that has need for such intermittent employment, are crucial in any country that has a chronic labor shortage or a tight labor market. In the United States the total volume of business in this industry is approximately $1.5 billion per year and, recently, has been increasing at the rate of 20% per year (Sloane, 1970). Cosgrove (1970) states that 2% of the workers in the United States are now in the temporary help field.

In light of this growth and importance, it is interesting to note that there have been few studies of this industry. In fact, a review of the literature suggests that only one major study has been published, namely, Moore's (1965) historical description of the development of the industry in the United States.

The present study is an analysis of some of the personal characteristics that are related to acceptance and refusal of employment among temporary help applicants. Specifically, the data pertain to the personal characteristics of two groups of temporary help applicants who were interviewed and hired by a temporary help firm during 1968. The first group consisted of applicants who, although they eventually left the firm, accepted employment during 1968. The second group was composed of applicants who were hired during 1968 but refused employment when the firm was able to place them.

There are at least two relevant reasons for such an analysis. First, it should be possible to reduce interviewing costs by pinpointing personal characteristics that are related to refusals of offers of employment. If such characteristics can be delineated, the firm can develop a short questionnaire and screen out those applicants not intrinsically interested in this employment, so as to eliminate the waste of valuable interviewing time. This procedure would be especially helpful to firms in this industry since they are bombarded with applicants. For example, Kelly Girls alone recruits 200,000 individuals each year (Cosgrove, 1970).

Second, the analysis of refusals should help in the evaluation of the directions and strategies that temporary help firms should pursue, especially since the number of refusals will probably increase as the labor market tightens and permanent jobs become available.

Method. The study was undertaken in a temporary help service company that has 21 offices located in New York, Connecticut, New Jersey, and Pennsylvania. Only the central office, however, which is located in New York City and is responsible for hiring within this locality, was used as the source of data in order to control for variations in labor market conditions.

The sample consisted of 299 individuals. Of the 299 individuals, 149 were temporary help applicants who accepted a job during 1968. The remaining 150 individuals were applicants who were interviewed during 1968, hired, but refused acceptance of a job. Both sets of applicants completed the employment interview.

Eighteen variables were examined relative to the two types of applicants, namely: age; years of school; sex; marital status; number of children; days available per week; expressed interest in a permanent job as a temporary employee; past experience in months; numbers of previous jobs; number of skills; type of job; current attendance at school; and the interviewer's perceptions as to the applicant's alertness, appearance, manner, attitude, voice, and grooming. The interviewer's perceptions were noted in terms of a 1–4 (excellent–poor) scale. Chi-square was employed as the methodological tool of analysis.

Results and discussion. Of the 18 variables, 5 were predictive of acceptance and refusal among applicants who were offered jobs by the temporary help firm. These were: days available per week, ex-

---

[1] The authors wish to thank Stephen J. Carroll, Jr. for his helpful criticisms and suggestions.

[2] Requests for reprints should be sent to Martin J. Gannon, Behavioral Science Division, Department of Business Administration, University of Maryland, College Park, Maryland 20742.

TABLE 1

PREDICTORS OF ACCEPTANCE AND REFUSAL OF JOB OFFER AMONG TEMPORARY HELP APPLICANTS

Variable	% Accepted job offer	% Refused job offer	N	χ^2
Type of job				
General secretary (lowest skill level)	81.0	19.0	21	
General clerk	50.3	49.7	185	
Typist	45.1	54.9	71	
Keypunch (highest skill level)	31.8	68.2	22	11.65**
No. skills				
0	38.8	61.2	49	
1	51.0	49.0	96	
2	41.1	58.9	73	
3	54.3	45.7	46	
4 and above	74.3	25.7	35	13.43**
Days available per week				
3 and under	65.9	34.1	41	
4–6	47.6	52.4	252	4.88*
Interested in permanent job as temporary employee				
Yes	56.2	43.7	144	
No	43.1	56.9	153	5.10*
Past experience (in mo.)				
60 and under	51.0	49.0	192	
61–80	43.5	56.5	23	
81–100	23.1	76.9	39	
Over 100	100.0	0.0	22	33.77***

Note.—$N = 299$ (N in each subtable does not total exactly 299 due to incomplete information.)

* $p \leq .05$.

** $p \leq .01$.

*** $p \leq .001$.

pressed interest in a permanent job with the firm, past experience in months prior to the employment interview, number of skills, and type of job.

As shown in Table 1, two measures of skill level were associated with the rate of acceptance, namely, type of job and number of skills. In both cases, an increase in skill level was related to a decline in the rate of acceptance. This finding is important in view of the fact that the temporary help service industry is currently expanding rapidly into areas of employment that require skilled and professional workers (Sloane, 1970). As the labor market tightens, skilled employees will generally find permanent jobs quickly. Hence, temporary help firms that specialize in skilled and professional services may very well find themselves in a position where they have few, if any, applicants who desire placement.

Another consideration in the acceptance of a job offer concerns the applicant's availability for placement. The applicants who indicated that they were available for placement more than 3 days a week tended to have a low rate of acceptance. Such a finding suggests that temporary help firms have difficulty in the procurement of workers who are available on a full-time basis. In view of the fact that business firms frequently desire and need temporary employees on a full-time basis for a short period of time, it is plausible that they may become disenchanted with temporary help agencies as sources of effective placements. This proposition is supported by the fact that agencies that place permanent workers generally are less effective than referrals by high schools, current employees, reemployment, and even walk-ins (Gannon, 1971). Further support for this proposition is noted in the fact that, as the applicant indicated his lack of desire for a permanent job as a temporary worker, his rate of refusal increased.

Finally, the type of individual who is attracted to the temporary help service industry is one who has the least or the most past experience measured in terms of months. Of the 22 applicants who had over 100 months' experience, not one refused the offer of employment. This finding suggests that the temporary help applicant who has a long history of past experience is serious in his desire to seek work as a temporary worker.

Conclusion. The analysis indicated that temporary help agencies may have difficulties in finding applicants who are available on a full-time basis, especially when the labor market tightens and/or the skill level is high.

The typical characterization of the temporary help industry has been very positive in the sense that its advantages are highlighted (Anonymous, 1963; Moore, 1965; Sloane, 1970). The present analysis, however, has demonstrated that many disadvantages and weaknesses, both actual and potential, have not been properly assessed. For example, temporary workers are most crucial when the labor market is tight, since most business firms are understaffed during this period. Yet, it is precisely during these periods that temporary help firms have much difficulty in providing reliable workers, especially in the skilled and professional areas.

Another potential difficulty with this type of service concerns the psychological set of the applicants who come to the temporary help firm. As indicated above, applicants who expressed a lack of interest in a permanent job as a temporary worker had a higher rate of refusal of job offers.

Finally, the type of individual who is attracted to this industry tends to be available for work only 3 days a week or less. Such a constricted amount of availability can become a source of difficulty, especially when the labor market is tight and business firms are interested in procuring workers on a full-time basis for a short period of time.

While further research into the subject of temporary help is obviously necessary, the present analysis has suggested particular difficulties and disadvantages that indicate that temporary help service is not a complete panacea for the ills of an imperfect labor market.

REFERENCES

COSGROVE, D. The temporary help service boom. *Personnel*, 1970, **47**, 44–48.

GANNON, M. Sources of referral and employee turnover. *Journal of Applied Psychology*, 1971, **55**, 226–228.

MOORE, M. The temporary help service industry: Historical development, operation, and scope. *Industrial and Labor Relations Review*, 1965, **18**, 554–569.

SLOANE, L. Part-time help: A permanent industry. *New York Times*, Business Section, January 4, 1970, 1, 26.

ANONYMOUS. People for rent: More companies use temporary office help provided by agencies. *Wall Street Journal*, December 12, 1963, 1.

(Received October 2, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 495–497

FACE-TO-FACE INTERACTION IN THE PEER-NOMINATION PROCESS

ARIE Y. LEWIN,[1] PETER DUBNO, AND WILLIAM G. AKULA[2]

Graduate School of Business Administration, New York University

In this note, the importance of face-to-face interaction in the peer-rating process was tested. Peer ratings were obtained from interacting group members participating in a management game and were compared with ratings obtained from observers who viewed 15-min. videotapes of these groups participating in a game decision-making session. There was no difference between the two sets of peer ratings leading to the conclusion that face-to-face interaction is not a critical variable in this judgmental process.

The validity of peer ratings in the prediction of various future performance criteria—for example, supervisors of life insurance agents (Weitz, 1958), managers (Roadman, 1964), and naval officers (Hollander, 1965)—has been shown to be remarkably high. Indeed, in a recent comprehensive review of the literature, Korman (1968) concluded that "peer ratings are better predictors than psychometric procedures [p. 319]" and better than most objective tests. Similar conclusions were reached by Lindzey and Byrne (1968) and Miner (1968). All of these writers, however, concluded by noting the lack of theory to explain the advantage of peer ratings and their validity in predicting future managerial performance or leadership.

Miner (1968), in commenting on the predictive potential of peer ratings over other selection or predictive methods, concluded that the explanation for this advantage at present was "conjectural." Lindzey and Byrne (1968), who conducted a survey of studies dealing with sociometric status (interchangeable with peer ratings), concluded that work in this area is almost totally devoid of theory. Furthermore, they concluded that after three decades of research on sociometric status, the typical investigation still is one in which regularly available measures are correlated with some index of choice status. Their criticism is particularly focused on the lack of a coherent theoretic structure to integrate the numerous findings to guide subsequent research toward critical questions. Without such a structure, work on sociometric status will remain as clear examples of limited returns to be expected from theoretic empiricism.

Finally, Korman (1968) acknowledged the explanatory problem by suggesting "systematic testing rather than a scattered, buckshot approach [p. 318]" as one kind of reorientation that was needed in predictive research.

Thus, it appears that in peer ratings we have a judgmental process that has high validity in the prediction of managerial performance but that at this time lacks a theoretical framework for satisfactory explanation of this process. In this note, the authors report on an experiment designed to test the importance of face-to-face interaction in the peer-rating process in an attempt to provide one small, but perhaps essential, portion of a theoretical formulation for peer ratings.

Statement of problem. The problem dealt with here is concerned with the well-known phenomenon that group members form judgments about one another as a result of interacting. Stated differently, it is the outcome of face-to-face social encounters that an individual interprets and applies to the judgmental process about the other participants of his group. Thus, Bales (1950) used face-to-face interaction as the means with which to define the basic concept and boundaries of any "small group." Bales thus assumes that interaction is a necessary ingredient for people to make judgments about others. That is, face-to-face interaction serves as a mediator to facilitate the learning process of humans and helps test their impressions or perceptions of each other. Similar assumptions about face-to-face interaction appear as a common thread in the studies of Homans (1950, 1961), Hare (1962), and Gibb (1969).

Gibb (1969) used "interaction" in both an abstract and physical sense. When referring to the simultaneous relations between independent variables, for example, Gibb is describing interaction in an abstract or quasi-mathematical form not relevant to the current study. When Gibb refers to (*a*) "interaction amongst constituent members," or (*b*) that members of a group should be seen, and see themselves, as having shared or compatible goals, and (*c*) that they be understood as satisfying individual needs through interactions with others, then he is referring to a physical or eye-to-eye interaction as a necessary part of the group process.

In this note, interaction is operationally defined as either social interaction or face-to-face interaction denoting the reciprocal influencing of the acts and feelings of persons. Consequently, the following general hypothesis is tested here, based on the Bales (1950), Hare (1962), and Gibb (1969) rationale: Assuming that face-to-face interaction is a determining factor in the process of individuals learning about each other within a small group, then peer ratings obtained from such an interacting group would be different from the ratings obtained about the same

---

[1] Requests for reprints should be sent to Arie Y. Lewin, Graduate School of Business Administration, New York University, 100 Trinity Place, New York, New York 10006.

[2] Now at the University of Bridgeport, Bridgeport, Connecticut.

TABLE 1

PROPORTIONS OF POOLED OBSERVER AGREEMENTS ON QUESTIONAIRE ITEMS FOR DIFFERENTLY RANKED INTERACTING GROUP MEMBERS

Item	Question 1: With whom can you work best?	Question 2: Who contributed most to achieving the goals of the team?	Question 3: Who contributed most to the analysis and solving of day-to-day problems?
Interacting group *S*s ranked first	.65	.69	.71
Interacting *S*s ranked either second, third, or fourth	.31	.31	.34
Interacting *S*s ranked last	.85	.87	.89
χ^2	15.0	11.3	9.3

Note.—$N = 97$ for four observer groups. Interacting group, $N = 14, df = 13$, *ns*.

group from a similar but *external* group of non-interacting observers.

Method. The research was conducted at the College of Business Administration of a large eastern university. A total of 95 evening students participated in a management game simulation. The business simulation used the JAN Lee's Executive Game for the IBM/1620 computer.

The 95 students were randomly allocated into 19 teams of five members each. All participants were exposed to a 1-hr. business game indoctrination and one practice session. Thereafter, they played three 20-min. sessions of the game.

The third 20-min. game session was recorded on videotapes. These recordings made up 15-min. samples of group decision making in the game. Although all 19 teams were videotaped, two teams were eliminated for viewing for external observers because of technical difficulties in the recording. Furthermore, three other teams had to be eliminated because they had one or more members absent at the peer-rating session. Consequently, 14 teams comprised the sample for the present study.

After having played the game for three 20-min. sessions, peer ratings were obtained from each participant with respect to his other group members in response to three questions covering the dimensions of: (*a*) With whom can you work best?, (*b*) Who contributed most to achieving the goals of the team?, and (*c*) Who contributed most to the analysis and solving of day-to-day problems? The sociometric questionnaire regarding these questions was adapted from Fiedler (1967).

The same questionnaire was then given to 97 external observers. These observers were asked to make the same peer ratings after observing 15-min. videotapes of each of the teams participating in the management game. In other words, the videotapes consisted of visual records of short time spans in the decision-making sessions of students in teams playing the game. The 97 external observers were divided into four groups of 20, 22, 25, and 30 members each. After a standardized practice session, each group was shown a randomly selected sequence of five management-game team videotapes. The 15-min. sequences of videotapes were shown one at a time and the observers were asked to make the peer ratings after each showing.

Analysis of results. The general hypothesis was tested within teams and across all *S*s. To obtain an overall impression for the degree of agreement between the peer nominations of the interacting groups and the external observer groups, the percentage of times observers in the four groups agreed with the ranked (1–5) interacting members was computed for each of the three criterion questions. The rankings were then pooled, according to the method outlined by Guilford (1954, pp. 181–188), to allow for a comparative analysis of the rankings obtained from the observer and interacting groups. Table 1 shows the pooled rankings of the observer groups for those interacting members ranked first and last; members ranked second, third, and fourth were combined into a single pooled ranking. The results presented show how the observers ranked the stimulus *S*s across the three criterion questions.

The obtained pooled rankings were then subjected to chi-square analysis; a theoretical frequency of 1.00 was employed for the first and fifth rank positions representing perfect similarity to the interacting groups' peer ratings. The middle ratings were removed from the analysis, however, because no sensible theoretical frequency value could be assigned. Indeed, the table shows that there was considerable disagreement between observers and interacters for these three positions, as might well be expected due to the well-known observation that *S*s in the middle ranges are often difficult to rank.

The extent to which the observers and interacting members' peer nominations agree can be tested by comparing the individual ranks that game participants received from their group members on each question with the ranks they received from the external observers. Another procedure employed for testing agreement of ratings involved the computation of a total rank for each group member by group from the peer ratings made by this group, and from the average of the external observers. Then, the difference between the two ranks was computed for each student on the 14 teams. The Wilcoxon Matched-Pair Signed Rank Test was used to determine whether these differences were significant. It will be recalled that the hypothesis stated a difference between the peer ratings obtained from interacting groups and from observers.

The Z values for the difference in ranks were .036, $-.068$, and $-.118$ for Questions 1, 2, and 3, respectively. The results reject the hypothesis that the differences in ranks are significant.

Discussion. In contrast to theoretical expectations, these results clearly demonstrate that face-to-face interaction does not appear to be a critical intervening variable in the judgmental process. It appears that limited vicarious interaction is sufficient for individuals to make judgments about other humans interacting in groups and that these judgments would be in agreement with those made by members of the interacting groups. Clearly, this raises the intriguing question that perhaps the judgmental process depends on the situation and is general to all people rather than being individualistic and complex. Furthermore, it suggests the additional possibility that viewers (our external observers watching the videotapes) engage spontaneously in a form of role playing that vicariously makes them members of the group they are watching for purposes of evaluating the real members of the group. This being the case, it would appear that "interaction" has psychological properties (social and projective) beyond those implied in the face-to-face notion, and that an explanation of the high level of predictive validity associated with peer ratings is far more complex than heretofore suggested.

Apparently, the face-to-face aspect of interaction is only one small physical ingredient in the process that does not even appear to be necessary for adequate peer ratings to occur. What seems to be much more probable is that some aspect of role behavior lies at the core of the judgment-interaction process. Some support for this point of view stems from Bem's (1965) replication of three dissonance experiments in which *S*s were prevented from experiencing the dissonance. He reported that his *S*'s verbal responses were essentially similar to those in the original dissonance experiments when they were given only a verbal description of the original dissonance situation. He concluded that *S* does not have to be physically in contact with a dissonance situation, or to actually experience any dissonance of his own, to respond similarly to the dissonance-generating treatments given to others. As a consequence of the findings presented here, further research employing role concepts may be warranted.

Furthermore, similarity in the ranking procedures of the internal group members and external observers may have been a function of their similarity with respect to values, attitudes, and culture. The external group members were selected from the same population of *S*s as the internal group members, that is, a population of business administration students. Further research should test the hypothesis with internal and external *S*s of different social, cultural, and even sexual characteristics.

REFERENCES

BALES, R. F. *Interaction process analysis: A method for the study of small groups.* Reading, Massachusetts: Addison-Wesley, 1950.

BEM, D. J. An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology*, 1965, **1**, 199–218.

FIEDLER, F. E. *A theory of leadership effectiveness.* New York: McGraw-Hill, 1967.

GIBB, C. A. Leadership. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology*. Vol. 4, Reading, Massachusetts: Addison-Wesley, 1969.

GUILFORD, J. P. *Psychometric methods.* New York: McGraw-Hill, 1954.

HARE, A. P. *Handbook of small group research.* New York: MacMillan, 1962.

HOLLANDER, E. P. Validity of peer nominations in producing a distant performance criterion. *Journal of Applied Psychology*, 1965, **49**, 434–438.

HOMANS, G. C. *The human group.* New York: Harcourt, Brace & World, 1950.

HOMANS, G. C. *Social behavior.* New York: Harcourt, Brace & World, 1961.

KORMAN, A. The prediction of managerial performance: A Review. *Personnel Psychology*, 1968, **21**, 295–322.

LINDZEY, G., & BYRNE, D. Measurement of social choice and interpersonal attractiveness. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology*. Vol. 2. Reading, Massachusetts: Addison-Wesley, 1968.

MINER, J. B. The early identification of managerial talent. *The Personnel and Guidance Journal*, 1968, **46**, 586–591.

ROADMAN, H. E. An industrial use of peer ratings. *Journal of Applied Psychology*, 1964, **48**, 211–214.

WEITZ, J. Selecting supervisors with peer ratings. *Personnel Psychology*, 1958, **11**, 25–35.

(Received September 17, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 5, 498–499

PERSONAL ORIENTATION INVENTORY AS A PREDICTOR OF SUCCESS IN PEACE CORPS TRAINING[1]

MICHAEL J. UHES[2] AND JOHN SHYBUT[3]

Carroll College *Northern Illinois University*

The present study investigated the relationship between scores on the Personal Orientation Inventory (POI) and success in Peace Corps training programs in Micronesia. The POI was administered to 92 Peace Corps trainees prior to the beginning of training and the results correlated with the final selection board ratings upon the completion of training. Significant relationships were found for the total group with considerable sex differences for several scales.

Attempts to predict trainee success in Peace Corps training programs and subsequent volunteer assignments have produced mixed results. Psychiatric interviews have been found to be ineffective (Fisher, Epstein, & Harris, 1967) as have "sickness"-oriented tests (Smith, 1965) and measures of intellectual ability (Dicken, 1969). Assuming that the crucial variables for Peace Corps trainee success could be best described in terms of positive mental health, Smith (1966) and Ezekiel (1968) found significant correlation coefficients ($r = .41$ and .34, respectively) between estimates of psychological effectiveness and volunteer success.

In recent years, Maslow (1962) has advanced the idea of the self-actualizing person as one who is developing and using all of his unique capabilities and potentialities. The Personal Orientation Inventory (POI) was developed by Shostrom (1966) to provide a measure of values and behavior seen to be of importance in self-actualization and positive mental health as opposed to psychopathology. These values and behaviors are similar to those described elsewhere as providing a coherent core of psychological attributes necessary for competent, effective Peace Corps functioning (Guthrie & McKendry, 1963; Smith, 1966).

The present note investigated the relationship between scores on the POI and trainee success in Peace Corps training programs in Micronesia.

Method. The *S*s for this study were 92 Peace Corps trainees (53 males, 39 females) participating in an in-country training program in Micronesia. The mean age of the sample was 22.2 yr. All but four of the trainees were college graduates. Each trainee had been screened for physical and psychiatric symptomatology prior to arriving at his respective training site.

The POI consists of 150 two-choice, comparative value and behavior judgments. The items were scored twice. First, for two basic scales of personal orientation, Inner Directedness (127 items) and Time Competence (23 items) and, second, for 10 subscales each of which measure a conceptually important element of self-actualization. Several studies have supported the construct validity (Fox, Knapp, & Michael, 1968; Graff, Bradshaw, Danish, Austin, & Altekruse, 1970; McClain, 1970; Shostrom, 1964) and the reliability (Illardi & May, 1968; Klavetter & Mogar, 1967) of the POI.

The criterion of success in the training programs was the Final Selection Board (FSB) rating. Upon completion of a 10-wk. training program, each trainee was assigned a composite rating of either 9 (upper quartile), 8 (third quartile), 7 (second quartile), or 6 (lower quartile) by the final selection board composed of the core training staff members. Each rating incorporated a trainee's performance in language training; technical studies; cross-cultural studies; and a general staff evaluation of interpersonal skills, motivation, and cultural adaptation. The composition of the final selection boards was not constant for all *S*s with relatively equal numbers being judged by four separate selection boards appropriate to the island-group to which the trainees were assigned. High interjudge agreement ($r = .87$) has been reported for FSB ratings (Goldberg, 1966), and such ratings have been found to be significantly related to a field performance criteria.

The POI was administered to all *S*s prior to arriving in Micronesia for training. Instructions at the time of administration indicated that the POI was a research instrument that would be scored on completion of the training program. Quadriserial r (Jaspen, 1946) was used to determine the magnitude of the relationship between POI scores and FSB ratings.

Results and discussion. The correlations between the POI and the FSB ratings are presented in Table 1. For the total sample, six of the scales reached the .05 level of significance or beyond. For male *S*s, 4 of the scales reached the .05 level or better; while for female *S*s, 7 of the 12 scales reached statistical significance. Of particular note are the correlation coefficients between the FSB ratings and two major scales, Time Competence and Inner Directedness. In-

---

[1] The data of this study were obtained with the assistance of the Peace Corps. This study, however, is the responsibiliy of the authors and should not be construed as an official Peace Corps document.

[2] Now at the Center for Cross-Cultural Training and Research, University of Hawaii, Hilo, Hawaii.

[3] Requests for reprints should be sent to John Shybut, Department of Psychology, Northern Illinois University, DeKalb, Illinois 60115.

TABLE 1

CORRELATIONS BETWEEN PERSONAL ORIENTATION INVENTORY SCALES AND FINAL SELECTION BOARD RATINGS

POI scale	Final selection board rating		
	Total r^a	Males r^b	Females r^c
Time Competence	.24*	−.20	.52***
Inner Directedness	.41***	.40**	.46**
Self-Actualizing Value	.24*	.33*	.12
Existentiality	.33**	.19	.48**
Feeling Reactivity	.13	.08	.18
Spontaneity	.06	.01	.32*
Self-Regard	.20	−.05	.33*
Self-Acceptance	.28**	.23	.37*
Nature of Man Constructive	.18	.28*	.02
Synergy	.13	.32*	.08
Acceptance of Aggression	.01	.06	−.14
Capacity for Intimate Contact	.28**	.26	.33*

Note.—POI = Personal Orientation Inventory.
[a] $N = 92$.
[b] $N = 53$.
[c] $N = 39$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

ner Directedness reflects the individual's autonomy, self-support, and, in general, freedom from rigid adherence to the social pressures and expectations. Time Competence may be viewed as an expression of use as opposed to misuse of time (Shostrom, 1966). Relatively high and significant correlations for Inner Directedness and FSB ratings suggest that this type of personal orientation favors adjustment in new situations, especially ones demanding personal initiative and relatively high degrees of self-support so characteristic of Peace Corps assignments.

The use of time, or Time Competence is also significantly related to the FSB ratings. An appreciable sex difference is noted for this scale. The difference is too great to be explained on the basis of higher scores reported for females in general (Shostrom, 1966). It is possible, however, that the difference may be largely due to the different concept of time and the male–female role designations in the Micronesian society as compared with the American culture. Thus, a transplanted American male who prides himself on being efficient in the use of time may be perceived by the Micronesians as an impatient individual and rated poorly on his adjustment to a relatively easy-going, nonschedulized daily approach to living. In contrast to males, behavior of females within the Micronesian culture seems to be more regulated, thus imposing on an outsider a clearly structured demand to conform to the existing sociotemporal orientation.

Of the remaining scales, Existentiality, one's flexibility in applying self-actualizing values and principles; Self-Acceptance, relative acceptance of one's self in spite of one's weaknesses and deficiencies; and Capacity for Intimate Contact, one's ability to develop meaningful relationships with others seem to be the most consistent across sexes in terms of validity. These personal orientations appear to be in line with the resourcefulness generally expected of the Peace Corps volunteers. With the exception of Feeling Reactivity and Acceptance of Aggression scales, which are not significant, the remaining five scales show at least one significant correlation with the FSB ratings. Lack of empirical data on which to base an explanation of the observed sex differences invites further exploration of this dimension.

While the need to test the effectiveness of the POI in identifying characteristics of successful performers in other populations is clear, these results tend to support Smith's (1966) hypothesis that indexes of psychological effectiveness (or positive mental health) provide the best predictors of competent Peace Corps functioning.

REFERENCES

DICKEN, C. Predicting the success of Peace Corps community development workers. *Journal of Consulting and Clinical Psychology*, 1969, **33**, 597–606.

EZEKIEL, R. S. The personal future and Peace Corps competence. *Journal of Personality and Social Psychology*, 1968, **8**(2, Pt. 2).

FISHER, J., EPSTEIN, L. J., & HARRIS, M. R. Validity of the psychiatric interview: Predicting the effectiveness of the first Peace Corps volunteers in Ghana. *Archives of General Psychiatry*, 1967, **17**, 744–750.

FOX, J., KNAPP, R. R., & MICHAEL, W. B. Assessment of self-actualization of psychiatric patients: Validity of the personal orientation inventory. *Educational and Psychological Measurement*, 1968, **28**, 565–569.

GOLDBERG, L. R. Reliability of Peace Corps selection boards: A study of interjudge agreement before and after board discussions. *Journal of Applied Psychology*, 1966, **50**, 400–408.

GRAFF, R. W., BRADSHAW, H. E., DANISH, S. J., AUSTIN, B. A., & ALTEKRUSE, M. The POI: A validity check. *Educational and Psychological Measurement*, 1970, **30**, 429–432.

GUTHRIE, G. M., & MCKENDRY, M. S. Interest patterns of Peace Corps volunteers in a teaching project. *Journal of Educational Psychology*, 1963, **54**, 261–267.

ILLARDI, R. L., & MAY, W. R. A reliability study of Shostrom's personal orientation inventory. *Journal of Humanistic Psychology*, 1968, 68–72.

JASPEN, N. Serial correlation. *Psychometrika*, 1946, **11**, 23–30.

KLAVETTER, R. E., & MOGAR, R. E. Stability and internal consistency of a measure of self-actualization. *Psychological Reports*, 1967, **21**, 422–424.

MASLOW, A. *Toward a psychology of being*. Princeton, N. J.: Van Nostrand, 1962.

MCCLAIN, E. W. Further validation of the Personal Orientation Inventory: Assessment of self-actualization of school counselors. *Journal of Consulting and Clinical Psychology*, 1970, **35**, 21–22.

SHOSTROM, E. L. A test for the measurement of self-actualization. *Educational and Psychological Measurement*, 1964, **24**, 207–218.

SHOSTROM, E. L. *Manual for the personal orientation inventory (POI). An inventory for the measurement of self-actualization*. San Diego: Educational and Industrial Testing Service, 1966.

SMITH, M. B. An analysis of "authoritarianism" among Peace Corps teachers. *Journal of Personality*, 1965, **33**, 513–535.

SMITH, M. B. Explorations in competence: A study of Peace Corps teachers in Ghana. *American Psychologist*, 1966, **21**, 555–566.

(Received September 14, 1970)

urnal of Applied Psychology
971, Vol. 55, No. 5, 500–502

CHARACTERISTICS OF VIOLATORS OF FORMAL COMPANY RULES

FRANS MULDER[1]

Technical University Twente, Enschede, The Netherlands

A comparison was made of employees having records of penalties for violating formal company rules with employees who were nonviolators. Differences in personality test scores were found. For example, violators scored lower on the factors cautiousness, vigor, responsibility, and emotional stability. Work groups that consider the leadership style of their supervisors to be social and well balanced rather than instrumental receive fewer punishment letters than groups having the opposite pattern of opinion about their supervisors.

Every organization makes rules that are considered necessary for its effectiveness and continued existence. Each employee agrees to meet certain standards of task performance, cooperation with other members of the organization, secrecy concerning company secrets, and respect for company property.

Wherever people work, however, deviations from these rules occur. Sometimes they occur to such a degree that it hampers the daily routine of the organization or threatens the well being of its members. Violators of formal company rules could be punished by means of a reprimand, a written warning, a transfer to another department, suspension, or discharge. Each year, at Hoogovens, an integrated iron and steel plant situated at Ijmuiden, the Netherlands, about 130 out of the 10,000 male Dutch employees receive a punishment letter (i.e., a written statement concerning a penalty). (Officials, women, Italian and Spanish workers were not included in this study.) In this article, violators of company rules are defined as people whose personnel record contains at least one punishment letter. Those workers whose record is "clean" are called nonviolators.

Sinha and Wherry (1965) have pointed out that despite the growing concern over increasingly frequent norm-violating behavior in the industrial world, very few systematic attempts have been made to investigate the determining or concomitant variables of deviants. Yet, "the tremendous number of reprimands, probations, temporary layoffs, demotions, and discharges that are invoked yearly by management indicate the perceived importance of discipline in business [Booker, 1969]."

The central theme in the present study is: What is the difference between violators and nonviolators in industry? Few studies have been published on this topic. Specifically, the author examined background, personality and ability differences, and characteristics of crews and their supervisors as possible sources of variance. Earlier work had suggested some personality correlates for possible study. For example, Cooper and Payne (1967) found correlations between extraversion and neuroticism on one side and the frequency of nonpermitted absence and frequency of lateness on the other side. In a Navy study, Knapp (1963, 1964) found the Socialization scale of the California Psychological Inventory and the Conformity scale of the Survey of Interpersonal Values correlated significantly with a delinquency criterion. Navy offenders were found to have attitudes favorable toward escapism and toward nonconformity to rules and regulations and were lower on socialization. Willingham and Ambler (1963) also found significant but low correlation between violating and the Gordon Personality Inventory Cautiousness and Vigor scales.

Method. The *S*s were as follows:

- Group A: 83 violators who received letters during 1961, 1962, and 1963;
- Group B: a group of 62 unpunished workers matched with Group A on the variables age, marital status, and length of service;
- Group C: 622 persons, hired in 1964, whose personnel record did not contain a punishment letter at the end of 1966;
- Group D: 125 persons, hired in 1964, whose personnel record did contain a punishment letter at the end of 1966 (Groups C + D equal the total of those hired in 1964);
- Group E: 266 violators and nonviolators who constituted 6 different crews;
- Group F: 19 psychopaths and 28 criminals.

For Groups A–E the following data were collected: the antecedents of punishment letters, if any such letters could be found in the company files; scores on the General Aptitude Test Battery (GATB); scores on the Kuder Vocational Preference Record; and scores on the Gordon Personal Profile and Inventory. The tests were administered during the first workday after hiring. The attitudes of Group E about their own group and their leader were collected during a survey not related to this project. Only the members of Group F were specially tested for this study.

Results. The violators and nonviolators could not be discriminated upon by age, percentage married and single, migration from other parts of the Netherlands, and length of service. The workers who were punished differed from those who were not punished on some variables. The violators were more likely to be divorced, to have very small (0 or 1 child) or very large families (5 or more children), to be more unskilled (fewer mechanical and electrical craftsmen, fewer operators and inspectors, more from the lower wage classes, more with an elementary school education), and to be less healthy (many at the medical examination before entering service declared that they suffered from diseases and mentioned a large variety of diseases).

---

[1] During the time this research was done, the author was psychologist at the Hoogovens plant, Ijmuiden, The Netherlands. Since January 1970 he has been in charge of research and education in industrial psychology at the Technical University Twente, Schubertlaan 6, Enschede, The Netherlands. Requests for reprints should be sent to this address.

No significant differences were found between Groups A and B for any of the GATB scores. From Groups C and D multiple-correlation and discriminant function analyses indicated that abilities did not discriminate violators from nonviolators.[2]

The results consistently confirmed only one hypothesis concerning the ability of violators: the standard deviation of the ability scores of violators tended to be significantly higher than that of nonviolators.

Violators scored significantly higher than nonviolators on the persuasive and literary factors and significantly lower on the factor measuring preference for mechanical activities of the Kuder Vocational Preference Record. For the scores on computational and artistic activities, the expected similarity between the groups was confirmed. The violators showed a lower preference for outdoor work and social service and a higher preference for musical jobs and office jobs. To summarize: norm violators in this industry prefer light jobs involving little physical exertion.

In agreement with findings of Cooper and Payne (1967), Knapp (1963, 1964, 1965), and Willingham and Ambler (1963) and with the fact that many authors reported inferiority feelings and bad adjustment in criminals, a factor analogous with a low score on the Gordon Personal Inventory, it was found that violators scored lower than nonviolators on cautiousness, vigor, emotional stability (on the Gordon Personal Profile), responsibility, and adjustment. For these factors, the product-moment correlation coefficients with violations were $-.46$, $-.26$, $-.36$, $-.44$, and $-.49$, respectively. A multiple-correlation coefficient of $-.59$ was obtained when predicting violations from both Gordon and Kuder variables.

The multiple-regression analysis was cross validated, but it proved impossible to discriminate the lightly and severely punished *Ss* using this technique. A discriminant function analysis proved to be more successful. It was also found that the more serious the violation, the lower were the scores on the personality traits relevant to norm violating. The scores of recidivists in light violating were similar to those of heavily punished first offenders and of criminal recidivists.

A Group Dimension Scale developed by Hemphill (1956) and adapted for the Netherlands by H. J. van Dongen (1969), was administered to the six crews included in Group E. The scale measures various aspects of group cohesion. The major hypothesis here was: The higher the scores on variables characterizing aspects of the cohesion in a group, the lower the number of punishment letters received by members of this group.

A group index was calculated from the sum of the group-cohesion variable scores divided by the average score for all groups. This resulted in a scale ranging from one to six. The rank correlation coefficient between punishment index (number of punishment letters of the group, divided by the number of members of the group) and this group index was rho $= -.83$ ($p < .01$).

A scale developed by H. J. van Dongen (1969) was used to measure the relation between leadership style and probability of violation. This scale measures social, instrumental, balanced, and dynamic aspects of leadership as seen by the group members. Social aspects include good relations with subordinates, tolerance to violation of rules, and applies more positive than negative sanctions. Instrumental aspects include giving priority to a quick working rate, high production, and strict observation of job instructions. Balance aspects include well-balanced, predictable, and stable characteristics. Dynamic aspects include readily receiving and implementing new ideas.

Two hypotheses were tested here: (*a*) the more social and well balanced a supervisor (as perceived by his subordinates), and (*b*) the less instrumental they think he is, the lower the punishment index of members of his group.

These hypotheses were confirmed in that crews with the lowest punishment letters per crew considered their bosses more social and well balanced and less instrumental than the crews with the highest punishment index.

The sum of the social and balance scores minus the instrumental scores divided by the mean scores of all the crews on the leadership variables is called here the leadership index. The rank order of this index correlated significantly with the punishment index (Violations), rho $= -.93$ ($p < .01$).

A multiple-regression analysis on the relative contribution of personality traits (P) and characteristics of groups (Gr) and leaders (L) gave the following results:

$$\text{Violations} = 1{,}714 - 4\text{P} - 4.6\ \text{Gr} - 7.2\ \text{L}$$

where:

Violations $=$ 1,000 × number of punishment letters per group divided by the number of group members,

P $=$ 100 × the result of substitution of scores in a formula with three Gordon

---

[2] Tables of mean scores on the different measures for the different groups have been deposited with the National Auxiliary Publications Service. Order Document No. 01558 from the National Auxiliary Publications Service of the American Society for Information Sciences, c/o CCM Information Science, Inc., 909 3rd Avenue, New York, New York 10022. Remit in advance $5.00 for photocopies or $2.00 for microfiche, and make checks payable to: Research and Microfilm Publications, Inc.

variables divided by the average score of all groups for this formula,
Gr = 100 × score on group cohesion scales divided by the average score of all groups, and
L = 100 × score on the leadership scales divided by the average score of all groups.

Discussion. It appears possible to discriminate norm violators from nonviolators from personality scores. The variance accounted for by personality, however, is low and may be of small practical utility. Leadership style and group cohesion appear to be more important variables than personality as predictors of norm violations. The correlation between leadership and punishment cannot be a result of selection of workmen by their leader because the correlation of the leadership index with the personality index is very low.

The correlation between group cohesion and norm violation is analogous with the results obtained by other investigators who studied delinquency. The results in these studies indicated that those parts of towns that are characterized by a low cohesion in community life showed delinquency rates eight to nine times higher than the cohesive parts of these communities. These results also suggest that norm violation is closely related to leadership style. Although norm violation and leadership style are measured independently, it is possible that the high correlation found here is partly caused by contamination because the boss judges his personnel and nominates violators for punishment. This comment, however, refers only to the "sanctions" aspects of social leadership style and not to other style aspects. There is, of course, a need to cross validate these suggestive results and replicate them in other industrial situations.

REFERENCES

BOOKER, G. S. Behavioral aspects of disciplinary action. *Personnel Journal*, 1969, **15**, 525–529.

COOPER, R., & PAYNE, R. Extraversion and some aspects of work behavior. *Personnel Psychology*, 1967, **20**, 45–57.

HEMPHILL, J. K. *Group dimensions, a manual for their measurement*. Columbus: Ohio State University, 1956.

KNAPP, R. R. Personality correlates of delinquency rate in a Navy sample. *Journal of Applied Psychology*, 1963, **47**, 68–71.

KNAPP, R. R. Value and personality differences between offenders and nonoffenders. *Journal of Applied Psychology*, 1964, **48**, 59–62.

KNAPP, R. R. Delinquency and objective personality factors. *Journal of Applied Psychology*, 1965, **49**, 8–10.

SINHA, J. B. P., & WHERRY, R. J., SR. Determinants of norm violating behavior in a simulated setting. *Personnel Psychology*, 1965, **18**, 403–412.

VAN DONGEN, H. J. *Sociaal-psychologische variabelen en het inzenden van ideeën. Microvariabelen en macrostructuren.* Stenfert Kroese, Leiden, 1969. (Social psychological variables and the suggestion of ideas. Microvariables and macrodimensions. With a summary in English.)

WILLINGHAM, W. W., & AMBLER, R. K. The relation of the Gordon Personal Inventory to several external criteria. *Journal of Consulting Psychology*, 1963, **27**, 460–472.

(Received April 23, 1970)

540

JOURNAL OF APPLIED PSYCHOLOGY

DECEMBER 1971 VOL. 55, NO. 6

Contents

ARTICLES

Pornography and the Sex Offender: Patterns of Previous Exposure and Arousal Effects of Pornographic Stimuli: *Royer F. Cook, Robert H. Fosen, and Asher Pacht* 503

Relationship between Task Difficulty and the Criterion: Should We Measure Early or Late?: *Edward L. Levine and Joseph Weitz* 512

Relative Influence of Job and Supervision on Shared Worker Attitudes: *Douglas B. Campbell* 521

Performance and Knowledge of Results as Determinants of Goal Setting: *L. L. Cummings, Donald P. Schwab, and Marc Rosen* 526

Effects of Contingent and Noncontingent Reward on the Relationship between Satisfaction and Task Performance: *David J. Cherrington, H. Joseph Reitz, and William E. Scott, Jr.* 531

Background Characteristics and Work Values Distinguishing Satisfaction Levels among Engineers: *Donald A. Wood* 537

Anxiety as a Function of Intrinsic-Extrinsic Job Orientation, the Presence or Absence of Observers, and Task Difficulty: *Shoukry D. Saleh* 543

Error and Variability in the Visual Processing of Dental Radiographs: *Irwin L. Goldstein and William H. Mobley* 549

Spatial and Temporal Information as Cues in a Time-Sharing Task: *Peter W. Dorfman and Irwin L. Goldstein* 554

Effects of Alcohol Ingestion on Tracking Performance during Angular Acceleration: *William E. Collins, David J. Schroeder, Richard D. Gilson, and Frederick E. Guedry, Jr.* 559

Influence of Expectation of Task Duration on Efficiency of Muscular Activity: *Stjepan Vidaček and Julius Wishner* 564

Price, Brand Name, and Product Composition Characteristics as Determinants of Perceived Quality: *Jacob Jacoby, Jerry C. Olson, and Rafael A. Haddock* 570

Effect of Fittingness, Type of Goods, and Type of Slogan on Brand Awareness: *Sasi Misra and Subhash Jain* 580

Characteristics of Nonrespondents among Workers: *Martin J. Gannon, Joseph C. Nothern, and Stephen J. Carroll, Jr.* 586

Effect of Personal Signature on the Initial Rate of Return of a Mailed Questionnaire: *Mary B. Kawash and Lawrence M. Aleamoni* 589

LIST OF MANUSCRIPTS ACCEPTED 558

This is the last issue of Volume 55.
Volume Title Page and Contents appear herein.

© 1971 by the American Psychological Association

INFORMATION FOR CONTRIBUTORS

Style. A professional article should possess certain characteristics: (*a*) conciseness and an apparent respect for reader time; (*b*) unambiguous and simple vocabulary with technical and erudite words used only when simpler ones would obviously be inadequate; (*c*) conformity to accepted technical style in tables, terminology, and references; (*d*) conclusions that are clearly related to the evidence presented. The reader should be led, step by step, from a statement of problem or purpose, through analysis of evidence, to conclusions and implications. Authors are encouraged to consult the little book entitled *The Elements of Style* by W. Strunk, Jr. and E. B. White (New York, Macmillan, 1962).

Criteria for evaluation. Manuscripts will be evaluated on the basis of several criteria, including: (*a*) significance in contributing new knowledge to the field, (*b*) technical adequacy, (*c*) appropriateness for the *Journal of Applied Psychology*, and (*d*) clarity of presentation.

Format. Manuscripts must be prepared in the format described in the *Publication Manual of the American Psychological Association* (1967 Revision), obtainable for $1.50 from the American Psychological Association, 1200 17th Street, N.W., Washington, D. C. 20036. Articles not prepared in this manner cannot be reviewed. Special attention should be given to the section on typing the manuscript (p. 48) and to the sections on tables, figures, and references. Note that *all copy must be double spaced*, including references, title, figure captions, etc. The senior author's last name should appear on each page in the upper left-hand corner. In reference lists, give journal titles in full; do not abbreviate.

Copies. All manuscripts must be submitted in triplicate. One copy should be the original typed copy, the other two, clear carbon copies or photo reproductions, only. Authors should check the final typing carefully and retain a copy of the manuscript as a precaution against loss in the mail.

Length. Before preparing a manuscript, the author should check several recent issues to get an idea of the approximate length of regular articles published in the *Journal of Applied Psychology*. (One printed page equals roughly three double-spaced typewritten manuscript pages.) A few longer articles of special significance may be printed from time to time as monographs. Occasionally the Journal will have a section of "Short Notes" featuring brief reports on studies which make some methodological contribution or constitute an important replication.

Abstracts. Each copy of the manuscript must be accompanied by an abstract of 100–120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts appeared in the *American Psychologist* (1961, **16**, 833), or they may be obtained from the Editor or from the APA Central Office.

Figures. The original drawing of a figure, or an 8 × 10-in. glossy print of the drawing, is needed for publication and must be submitted with the original typed copy of the manuscript. Duplicate copies of the figure may be photographic or pencil-drawn copies. Figures must be hand-lettered; typewritten lettering is not acceptable.

Early publication. Regular articles are published as nearly as possible in the order in which they were received. Although everything possible is done to keep publication lag to a minimum, at present there is a 10–12-mo. period between the time of receipt and publication of an article. The author of an accepted manuscript may secure early publication, usually in the first issue to go to press after its acceptance, by arranging for the article to be printed as extra pages, increasing the number of pages that subscribers receive. Information about charges for early publication will be sent with the letter of acceptance.

Reprints. No gratis reprints are supplied. Reprints may be ordered from the printer prior to publication.

Copyrighted material. The following policies govern reprinting or other forms of reproduction of materials copyrighted by the American Psychological Association: Written permission must be obtained for copying or reprinting any tables, figures, or text of more than 500 words in length. Persons or organizations who obtain permission to reproduce any such copyrighted materials must include on the first page of the reproduced materials the exact copyright notice which appeared on the Association's original publication. There will be a charge of $10.00 per page for such permission except when the requester is the author of the article to be reprinted, or when the article is to be reproduced in a limited number of copies solely for instructional purposes.

Journal of Applied Psychology
1971, Vol. 55, No. 6, 503–511

PORNOGRAPHY AND THE SEX OFFENDER:

PATTERNS OF PREVIOUS EXPOSURE AND AROUSAL EFFECTS OF PORNOGRAPHIC STIMULI[1]

ROYER F. COOK,[2] ROBERT H. FOSEN[3]

American Institutes for Research, Washington, D. C.

AND ASHER PACHT

Wisconsin Division of Corrections

One hundred twenty-nine inmates (63 sex offenders, 66 criminal code offenders) were shown a series of 26 slides depicting different sexual behaviors. The *S*s were also interviewed about their past exposure to pornography. No differences were found between the two groups on the measure of rated overall sexual arousal to the slides. Numerous differences were found between the two groups in their past exposure to pornography; sex offenders generally experienced less frequent and milder exposure to pornography than did the criminal code offenders. Factor analyses of arousal ratings to the different slides identified several arousal factors and suggested possible differential relations to type of sex offense.

With the current proliferation of depictions of sexual behavior in books, magazines, films, and theater, a larger segment of society appears to be engaging in vicarious sexual experience—viewing and reading about sexual behavior. The trend has been accompanied by renewed interest on the part of both psychology and law in the relationship between exposure to pornography and the commission of a sexual offense (Cairns, 1971).

Only during the past decade have researchers begun systematically to investigate this relationship, typically employing survey-interview methods to determine the type and frequency of exposure to pornography (e.g., Gebhard, Gagnon, Pomeroy, & Christenson, 1965; Thorne & Haupt, 1966). These studies compared the exposure experiences of sex offenders to those of nonsex offenders and/or groups from society at large. Results have been largely negative, indicating that exposure rates of sex offenders have not significantly differed from their control groups.

More recently, several survey-interview studies supported by the Commission on Obscenity and Pornography have generally shown that sex offenders had significantly less exposure to pornography as children and adults than comparable control groups (Goldstein, Kant, Judd, Rice, & Green, 1970; Johnson, Kupperstein, & Peters, 1970; Walker, 1970). However, none of these studies has measured the immediate effects of pornographic material on the arousal level of sex offenders.

The present research was designed to identify differences in immediate and subsequent arousal levels between individuals in prison for sex crimes and those individuals in prison for crimes unrelated to sexual behavior when these individuals are exposed to pornographic stimuli depicting different types of sexual behavior. Also investigated were differences between these groups in the type and frequency of the individual's prior history of exposure to pornography: (*a*) during preadolescence and adolescence, and (*b*) more immediately prior to his crime. Finally, an exploratory study was performed on the existence of general or specific arousal patterns to different types of pornographic stimuli and possible relationships of these more specific patterns to sexual offenses.

---

[1] This study was supported by a grant from the Federal Commission on Obscenity and Pornography. The authors would like to thank Albert S. Glickman for his helpful guidance throughout the research and Raymond Anderson for his assistance in collecting the data.

[2] Requests for reprints should be sent to Royer F. Cook, who is now affiliated with the Department of the Army, Motivation and Training Laboratory, Commonwealth Building, 1300 Wilson Boulevard, Arlington, Virginia 22209.

[3] Now Assistant Commissioner, New York Department of Correctional Services, Albany, New York.

METHOD

Subjects

One hundred twenty-nine male inmates at the Wisconsin State Prison served as *S*s. All *S*s were told beforehand that participation was voluntary and that their responses were confidential. Eighteen inmates refused to participate; they were rather evenly distributed across offense types. The most prominent reason for refusal was an unwillingness to cooperate with the institutional authorities. Despite assurances, a few of the sex offenders felt that the experiment might have a bearing on their chances for parole. The *S*s were 66 persons convicted of criminal code offenses unrelated to sexual behavior (CC) and 63 sex offenders (SO).

The CC sample was representatively selected from a population of 850 inmates; the SO sample from a population of 114. The CC sample contained inmates convicted of burglary ($N = 17$), homicide ($N = 18$), forgery ($N = 14$), and robbery ($N = 17$). The SO sample contained inmates convicted of indecent liberties with females ($N = 21$), indecent liberties with males ($N = 16$), attempted rape ($N = 10$), and rape ($N = 16$). The two samples were reasonably comparable in average education (SOs = 8.7 yr., CCs = 8.9 yr); average general abilities (SOs = 89.7 on Cattell Culture Fair Test, CCs = 92.2 on Cattell Culture Fair Test), and average age (SOs = 28.6 yr., CCs = 29.7 yr.). The only practical restrictions placed on selection were that all *S*s were to be functionally literate and that only the most recently incarcerated were chosen to minimize deprivation effects. Thus, the large majority of *S*s in both samples had been incarcerated for less than 3 yr.

Apparatus and Materials

The stimulus materials were 26 color slides taken from four 10-min. films. Table 1 describes the thematic content of the slides. A Kodak 850 Carousel projector automatically presented a 22½ in. × 15½ in. image on a screen approximately 7 ft. from *S*. Data were collected in one of two rooms located in the clinical services area of the prison.

Pretests

A pretest was conducted with seven staff members of the American Institutes for Research to try out the rating procedures and to provide a basis for reducing the number of slides from 49 to 26. A second pretest of instruments and procedures was conducted at the Wisconsin Correctional Institution using two CCs and two SOs.

Design

Three measures[4] were used: (*a*) a set of ratings by means of which *S* rated his level of sexual arousal to each slide as each was presented, (*b*) a questionnaire eliciting *S*'s responses to 22 questions about his earlier experiences with pornography, and (*c*) a questionnaire given to *S* the morning after the experimental session.

With regard to the immediate arousal effects of the slides, the two variables of major interest were offender type and presentation order (story line or random). In the story-line presentation the slides were arranged in a realistic sequence from fully clothed, through petting, to intercourse; in the random presentation, *S* received the 26 slides in one of three randomized sequences. These variables (two presentation orders and eight types of offenses) were arranged in a 2 × 8 factorial design to permit an independent-groups analysis of variance on the ratings data (Winer, 1962).

With respect to the pornographic experience questionnaire, the 22 questions were categorized as follows: (*a*) type and frequency of *S*'s experience with pornography, during year before imprisonment, (*b*) type and frequency of *S*'s experience with pornography during preadolescence and adolescence, (*c*) *S*'s view of the educational function of pornography, (*d*) *S*'s perception of the adjustment function of pornography, (*e*) *S*'s motivations for experiencing pornography, and (*f*) *S*'s view of the immediate effects of pornography on criminal behavior.

The questionnaire administered the morning after the experimental session asked the following five questions:

1. How often did you think about the slides after you left the experiment yesterday?
2. When you left the experiment yesterday, how excited did you feel?
3. When you left the experiment yesterday, did you feel "turned off," disgusted?
4. Did you have any sexual fantasies after the experiment (daydreams, etc.)?
5. Did you engage in any sexual behavior yesterday or last night (e.g., masturbated)?

Procedure

The experimental session consisted of two parts. During the first half of the session, *S* viewed the slides presented at 15-sec. intervals and rated his level of sexual arousal to each slide. In the second part, *E* interviewed *S* about his preprison experience with pornography using the questionnaire described above.

The *S* reported to the clinical services area and was escorted to the experimental room by one of the two *E*s. He was told the following:

> The Commission on Obscenity and Pornography has asked us, a private research institute, to do a study of reactions to pornography among inmates. We simply want to find out how people react to pornography, pictures of sexual behavior. You will look at some pornographic slides and make some judgments about them. We would like you to serve as a subject. Read over this consent form carefully, and if you would like to participate, sign the form at the bottom.

After filling out the mood descriptor, *S* was told that he would see a series of 26 slides which he should rate according to a 6-point scale (1 = neutral, not exciting; 2 = slightly exciting; 3 = moderately exciting; 4 = very exciting; 5 = highly exciting; 6 = extremely exciting). It was emphasized that he was to rate the slides according to *his* level of arousal, not as he thought someone else might rate them. The *S* was shown five preview slides with a representative range of arousal values (as obtained in the pretests) to establish a frame of reference, then was given a pad with a separate page for each slide rating. The *E* set the projector so that each slide would be presented automatically for 15 sec. The *E* then left the room while *S* viewed and rated the slides. The *E* returned after the last slide was presented, gave *S* another mood descriptor to fill out, then began the exposure interview.

---

[4] A fourth measure, a mood descriptor given to *S* before and after slide viewing, is not presented here. Many *S*s had difficulty understanding the scales and many of the data appeared invalid.

RESULTS

Immediate Arousal Effects

Mean ratings of individual slides for all *S*s are given in Table 1. In terms of thematic content, most highly rated are those slides showing intercourse in which the female is over the male. Next are the nude petting scenes, followed by cunnilingus, then fellatio slides. Toward the bottom of the ranking are the partially clothed petting and undressing scenes.

TABLE 1

MEAN AROUSAL RATINGS OF SLIDES

Content	X	SD
1. Intercourse, ventral-ventral, F over M	3.26	1.63
2. Intercourse, ventral-dorsal F over M	3.23	1.63
3. Nude petting, M hand on F genitals	3.14	1.65
4. Nude petting, M hand on F genitals, kissing breast	3.04	1.44
5. Intercourse, ventral-dorsal, F over M	3.02	1.61
6. F masturbating	3.01	1.57
7. M approaching cunnilingus	2.90	1.62
8. Cunnilingus	2.89	1.73
9a. Cunnilingus	2.86	1.72
9b. M and F in underclothes, hands on genitals	2.86	1.72
10. Fellatio, M over F	2.85	1.66
11. Fellatio, F over M	2.84	1.72
12. Intercourse, ventral-dorsal, each on side	2.83	1.56
13. M over F, genitals on breasts	2.75	1.56
14. Intercourse, ventral-ventral, M over F	2.73	1.50
15. Fellatio, M over F	2.65	1.62
16. Nude F reclined on back	2.61	1.35
17. Petting, partially clothed, F hand on M genitals	2.60	1.36
18. F partially clothed, undressing M	2.37	1.26
19. F partially clothed, undressing M	2.19	1.30
20. F and M undressing each other	2.13	1.21
21. F and M with blanket over them	1.83	1.14
22. F and M dressed, about to kiss	1.66	1.00
23. F and M dressed, about to kiss	1.47	.81
24. Nude M	1.29	.82
25. Fully clothed couple	1.12	.43

Note.—Abbreviations: F = female, M = male.

Mean rankings by SOs and CCs were nearly identical. The *t* tests for individual slides showed significant differences between SOs and CCs on only two of the slides (SOs having higher ratings), both of which were relatively low in arousal level (Slide 21, $t = 2.51$, $df = 127$, $p < .01$; Slide 25, $t = 1.97$, $df = 127$, $p < .05$).

TABLE 2

MEANS AND STANDARD DEVIATIONS OF AVERAGE AROUSAL RATINGS

Offense	Presentation		
	Story	Random	Total
Indecent liberties with females			
X	2.32	2.06	2.18
SD	1.28	1.53	1.27
Indecent liberties with males			
X	2.62	2.54	2.57
SD	1.59	1.52	1.55
Attempted rape			
X	3.28	2.91	3.10
SD	1.51	1.43	1.48
Rape			
X	3.08	2.75	2.94
SD	1.57	1.63	1.60
Burglary			
X	2.64	2.49	2.57
SD	1.60	1.63	1.61
Homicide			
X	2.52	2.78	2.63
SD	1.68	1.54	1.62
Forgery			
X	2.49	1.93	2.21
SD	1.52	1.20	1.40
Robbery			
X	2.03	2.88	2.43
SD	1.33	1.76	1.60

Table 2 displays the means and standard deviations of overall arousal ratings (average across all slides), categorized according to method of presentation (story line or random) and type of offense. An analysis of variance revealed no significant effect of either preentation ($F = .15$ $df = 1{,}113$) or offense $F = 1.40$ $df = 7{,}113$). The Presentation Offense interaction was also nonsignificant $F = .77$ $df = 7{,}113$). The overwhelming im-ression is one of similarity rather than dif-erences between overall arousal ratings of Os and criminals who are not sex offenders.

revious Exposure Patterns

Most of the data were analyzed by chi-quare. With 2 × 2 contingency tables with above 40, Yates' Correction for Continuity was used. Fisher's Exact Test was used on × 2 tables with $N < 20$ (Siegel, 1956). Most f the data are simple counts of particular questionnaire item responses. The one excep-ion was the use of a combined severity-frequency score. This was an effort to obtain an ndex of intensity of exposure that is some function of frequency and degree of arousal value of the material. The severity-frequency score was derived by first assigning ratings according to the severity of the material. A rating of 1 was assigned if *S* had viewed no pornography. A 2 rating was given for exposure to live entertainment (burlesque girls, go-go dancers), erotic novels, semierotic movies shown in regular theaters, and magazines displaying seminude males or females (e.g., *Playboy*). A rating of 3 indicated exposure to hard-core pornography: photographs, drawings, or films (stag films, "smokers") depicting explicit sex acts. The *S*'s severity-frequency score was obtained by adding up his ratings. Therefore, if *S* had seen a stag film, seen go-go dancers, and read *Playboy*, his severity-frequency score would have been 7.

TABLE 3

Experience with Pornography during Preadolescence

Type	Offense*	
	Sex offender	Criminal code violator
None	43	31
Occasional magazines	6	10
Hard-core pornography	11	22

* $p < .05$.

TABLE 4

Experience with Pornography during Adolescence

Severity frequency	Offense*	
	Sex offender	Criminal code violator
Infrequent or mild	21	10
Frequent or hard core	33	46

* $p < .05$.

No significant differences were found between SOs and CCs on the severity-frequency index during the year just prior to imprisonment.

As Table 3 shows, during preadolescence (6–10 years) there are significant differences between SOs and CCs with regard to type of exposure; the major result here is that SOs report *less* exposure than CCs.

These differences carry through to adolescence (see Table 4): SOs experience milder forms of pornography and at a less frequent rate than do CCs.

When *S* was asked specifically how often he was exposed to pornography during his childhood, again SOs reported significantly less exposure than CCs ($\chi^2 = 11.92$, $p < .05$).

While there were no differences between SOs and CCs as to whether they feel they learned anything about sex from pornography, it should be noted that 49 *S*s felt they did learn something about sex, while 33 *S*s said that they learned information that was misleading.

Most *S*s in both groups did not feel that they would have been better off sexually if they had been exposed to more pornography during their youth. But additional responses to this question show that more CCs ($N = 6$) than SOs ($N = 1$) said that they had "plenty available" as youngsters, while more SOs ($N = 11$) than CCs ($N = 2$) said they would have preferred "some sex education." Although our *N*s here are small, these differences were significant when analyzed by Fisher's Exact Test ($p < .02$).

There are significant differences between SOs and CCs in the *kind* of sexual behavior reported after viewing pornography (see Table 5). The CCs said they typically engaged in sexual relations after viewing pornography, while SOs more frequently reported masturbation.

Most *Ss* claimed not to have viewed pornography during the 24 hr. before committing the offense. When asked if viewing pornography encouraged them to commit their offense, only four SOs and no CCs answered affirmatively.

Post Slide-Viewing Sexual Behavior

The follow-up questionnaire yielded no significant differences between SOs and the nonsex offenders. Most *Ss* (90%) claimed not to have been "very excited" upon leaving the experiment, nor did they feel disgusted upon leaving the experiment (87%). Most claimed not to have sexual fantasies (82%) or to have engaged in sexual behavior afterward (90%).

Factor Analyses of Arousal Ratings

In an exploratory attempt to identify the structure of the erotic response to visual depictions of sexual behavior, the intercorrelations of arousal ratings were factor analyzed employing ratings of the 26 slides as variables. Although the analysis of variance revealed no differences in total arousal levels between the two methods of presentation (random and story-line sequence) it was conjectured that the factor structures might be different. Consequently, two separate factor analyses were carried out, one for the group receiving the slides in a random sequence and one for the group receiving the slides in a story-line sequence.

TABLE 5

REPORTED SEXUAL BEHAVIOR AFTER VIEWING PORNOGRAPHY

Behavior	Offense*	
	Sex offender	Criminal code violator
Masturbation	16	9
Sexual relations	6	11
Both masturbation and sexual relations	4	13

* $p < .05$.

Because *N*s were relatively small, 66 for story-line and 63 for random sequences, the resulting factor structures must be considered tentative. Similarly, the subsequent comparison of group factor means is intended mainly to suggest directions for further research.

For each analysis the squared multiple correlations were used as communality estimates. The squared multiple R is generally considered a conservative communality estimate, the "true" communality lying somewhere between the squared multiple R and 1. Oblique rotations were used. Table 6 presents the complete rotated factor matrices for each analysis.

The two presentation orders did result in different factor structures. While the story-line presentation yielded two larger factors and one smaller factor, the random presentation produced a more balanced, differentiated structure (see Table 6).

The factor interpretations for the random presentation are presented first (see Table 7). Table 7 presents, for each factor, the slides with the highest loadings; also presented are the rankings of the slides on "arousal ratings" (from Table 1). Only loadings above .40 are presented.

The first factor from the random presentation yielded high loadings on slides involving the *female genitalia*.

The second factor, *foreplay*, loaded heavily on the slides depicting petting and foreplay. In contrast to the female genitalia factor, the slides loading heavily on this factor were all ranked rather low (17–23) in arousal value.

The third factor was identified as a *fellatio* factor, common for the only three slides depicting this kind of sexual behavior. These slides ranked in the middle range of arousal values obtained.

The final factor extracted was the *intercourse* factor, loading heavily on slides depicting various positions of intercourse. Slide 14 (intercourse, ventral-ventral, M over F) had a loading of .39.

With respect to the story-line presentation of stimuli, Table 8 presents, for each factor, the slides with the highest loadings. Table 8 also

TABLE 6

FACTOR LOADINGS AND COMMUNALITIES

Content	Arousal rank	Random factors					Story-line factors			
		1	2	3	4	h²	1	2	3	h²
1. F and M clothed	25	.26	.35	.27	.10	.25	.01	.03	.74	.60
2. Dressed petting	23	.03	.80	.03	.34	.70	.00	.24	.73	.86
3. Dressed petting	22	.06	.72	.07	.10	.62	.11	.52	.46	.78
4. F and M undressing	20	.00	.69	.02	.08	.73	.05	.61	.23	.74
5. F and M undressing	18	.00	.36	.15	.39	.56	.02	.62	.25	.89
6. F and M undressing	19	.19	.57	.03	.34	.72	.06	.67	.08	.78
7. Petting	9b	.14	.27	.16	.30	.73	.20	.57	.52	.85
8. Petting	17	.16	.43	.25	.34	.72	.15	.58	.01	.74
9. Nude F	16	.44	.14	.08	.11	.58	.18	.51	.04	.69
10. Nude M	24	.05	.28	.30	.42	.32	.24	.62	.25	.41
11. F masturbating	6	.66	.06	.04	.02	.80	.65	.06	.03	.75
12. Nude petting	3	.40	.01	.01	.29	.74	.56	.26	.04	.85
13. Genitals on breasts	13	.10	.19	.36	.10	.53	.60	.24	.17	.85
14. Fellatio	15	.03	.06	.71	.04	.77	.72	.02	.03	.84
15. Fellatio	10	.05	.05	.72	.11	.89	.73	.03	.05	.78
16. Fellatio	11	.20	.03	.53	.02	.73	.72	.01	.05	.79
17. Nude petting	4	.25	.25	.02	.31	.76	.58	.09	.13	.72
18. Cunnilingus	7	.50	.06	.03	.14	.63	.57	.16	.11	.80
19. Cunnilingus	8	.63	.13	.14	.08	.79	.74	.10	.10	.78
20. Cunnilingus	9a	.64	.03	.10	.06	.85	.75	.11	.05	.76
21. Intercourse	1	.06	.05	.07	.61	.77	.73	.02	.06	.85
22. Intercourse	5	.14	.09	.22	.44	.67	.77	.03	.01	.87
23. Intercourse	2	.38	.03	.01	.28	.68	.72	.07	.10	.88
24. Intercourse	12	.22	.07	.08	.48	.77	.74	.07	.02	.77
25. Intercourse	14	.10	.08	.00	.39	.45	.41	.27	.15	.53
26. F and M under blanket	21	.00	.66	.21	.13	.65	.09	.50	.06	.35

Note.—Slides are listed according to the presentation order of the story-line sequence. Abbreviations: F = female, M = male.

presents the order of presentation along with the arousal rankings and loadings over .40. The first factor is general to *heavy petting and intercourse* slides which progress in content from nude petting, through oral-genital contact, to intercourse. These slides were among the most erotic (13 out of the highest 15 on arousal ratings) and appeared toward the end of the sequence (11–15, none earlier).

The second factor is a *foreplay* factor which yielded large loadings on the slides showing petting and foreplay activities. The foreplay slides were ranked next lowest in arousal level and were intermediate in the story sequenee (all but one of the slides between third and tenth in the sequence).

The final factor extracted was labeled an *anticipation* factor, consisting of the first set of slides in the story sequence. This story-line presentation tended to differentiate the anticipation factor from the foreplay factor, but the random sequence did not.

Differentiation of Offenders by Factor

Our original analyses, comparing different groups of offenders, were based on total scores summed across all 26 slides. Would more differentiated results be obtained if we examined the specific arousal patterns identified by the factor analyses? On the basis of the factor analyses, mean arousal ratings of each offense group were computed separately for those slides which loaded heavily on each separate factor. Thus, four arousal scores, representing total arousal ratings on the slides within each random factor, were calculated for each offender. Similarly, three arousal scores from each story-line factor were computed for each offender.

TABLE 7

AROUSAL RANKS AND HIGHEST FACTOR LOADINGS FOR SLIDES PRESENTED IN RANDOM ORDER

Factor	Arousal rank	Content	Loading
Female genitalia	6	F masturbating	.66
	9a	Cunnilingus	.64
	8	Cunnilingus	.63
	7	M approaching cunnilingus	.50
	16	Nude F reclined on back	.44
	3	Nude petting, M hand on F genitals	.40
Foreplay	23	F and M dressed, about to kiss	.80
	22	F and M dressed, about to kiss	.72
	20	F and M undressing each other	.69
	21	F and M under blanket	.66
	19	F partially clothed undressing M	.57
	17	Petting, partially clothed, F hand on M genitals	.43
Fellatio	10	Fellatio, M over F	.72
	15	Fellatio, M over F	.71
	11	Fellatio, F over M	.53
Intercourse	1	Intercourse, ventral-ventral, F over M	.61
	12	Intercourse, ventral-dorsal, each on side	.48
	5	Intercourse, ventral-dorsal, F over M	.44
	24	Nude M	.42

Note.—Abbreviations: M = male, F = female.

While these results must be regarded as preliminary and in need of cross-validation, several findings may be of particular interest.

The results indicate that SOs appear to have a flatter profile of means across the four factors than do the CCs. In particular, the SOs convicted of indecent liberties with females had significantly lower means on the fellatio factor slides than either the other SOs ($t = 1.98$, $df = 30$, $p < .05$) or the CCs ($t = 2.25$, $df = 40$, $p < .05$).

For the story-line group, it was found that the scores on the heavy petting and intercourse factor were significantly higher for rapists (those convicted of rape and attempted rape) than for the other SOs ($t = 2.10$, $df = 29$, $p < .05$) and for the CCs ($t = 2.14$, $df = 47$, $p < .05$).

DISCUSSION

The overall immediate arousal effects of exposure to the pornographic stimuli were generally characterized by a lack of differences among the offense groups. Both of the major samples, SOs and CCs, exhibited relatively low overall arousal, as compared to that of college students who viewed highly similar erotic material (Brady & Levitt, 1965). On the other hand, the rankings of slides by both SOs and CCs closely parallel the rankings by the college students. A large majority of all *S*s reported little change in behavior during the 24 hr. after viewing the material.

Numerous differences were found between the two major samples in their previous exposure to pornography. In contrast to the CC, the SO had less frequent exposure to pornography during preadolescence and adolescence and, when exposed, it was usually to milder forms of pornography. These findings are compatible with other recent evidence (Goldstein et al., 1970; Johnson et al., 1970) and suggest that childhood experiences that encourage sexual repression are associated with antisocial sexual behavior. In short, these data indicate a negative relationship between exposure to pornography and the tendency to commit a sex crime.

The exploratory factor analyses revealed a number of interesting points about erotic responses to pornography. The identification of independent factors of erotic response indicates that there are marked differences in arousal patterns to depictions of different forms of sexual behavior and that these patterns cluster into a limited number of identifiable factors. Thus, for most purposes it is less meaningful to talk about overall

TABLE 8

ORDER OF PRESENTATION, AROUSAL RANKINGS, AND HIGHEST FACTOR LOADINGS OF STORY-LINE SLIDE PRESENTATION

Factor	Presentation order	Arousal rank	Content	Loading
Heavy petting and intercourse	22	5	Intercourse, ventral-dorsal, F over M	.77
	20	9a	Cunnilingus	.75
	19	8	Cunnilingus	.74
	24	12	Intercourse, ventral-dorsal each on side	.74
	15	10	Fellatio, M over F	.73
	21	1	Intercourse, ventral-ventral, F over M	.73
	14	15	Fellatio, M over F	.72
	23	2	Intercourse, ventral-dorsal, F over M	.72
	16	11	Fellatio, F over M	.72
	11	6	F masturbating	.65
	13	13	M over F, genitals on breasts	.60
	17	4	Nude petting, M hand on F genitals, kissing breast	.58
	18	7	M approaching cunnilingus	.57
	12	3	Nude petting, M hand on F genitals	.56
	25	14	Intercourse, M over F, ventral-ventral	.41
Foreplay	6	19	F partially clothed, undressing M	.67
	5	18	F partially clothed, undressing M	.62
	10	24	Nude M	.62
	4	20	F and M undressing each other	.61
	8	17	Petting, partially clothed, F hand on M genitals	.58
	7	9b	M on F in underclothes, hands on genitals	.57
	3	22	F and M dressed, about to kiss	.52
	9	16	Nude F reclined on back	.51
	26	21	F and M with blanket over them	.50
Anticipation	1	25	Fully clothed couple	.74
	2	23	F and M dressed, about to kiss	.73
	7	9b	M and F in underclothes, hands on genitals	.52
	3	22	F and M dressed, about to kiss	.46

Note.—Abbreviations: M = male, F = female.

arousal to pornographic stimuli than it is to talk about more specific patterns. The identification of these factors suggest the extent as well as the limits of these patterns. For example, arousal to female genitalia seems a separable pattern but is general to scenes involving nude females at rest, female masturbation, nude petting, or cunnilingus. And individuals aroused by depictions defined by this factor are not necessarily the same individuals as those aroused by depictions defined by the other factors of intercourse, fellatio, or foreplay.

The materials used in the present study are relatively limited. Future work on the structure of erotic responses should employ a larger variety of stimuli (in terms of media as well as content) and additional response measures (e.g., physiological).

The finding of a somewhat different factor structure with a random sequence, as compared with a story-line sequence, has a number of implications. The random sequences produced a much more differentiated factor matrix, separate factors having been extracted for each of the major sexual themes presented in the slides. With the more realistic story-line sequence, the response seems to follow a more integrated course, and the factors extracted are shaped according to sequence as well as theme. Story-line sequence probably involves a greater amount of vicarious experiencing with the entire act, with *S* viewing them as a realistic integrated flow of events. Yet different individuals appear differentially aroused by factors at different stages of the sequence. The more fragmented random presentation probably lessens *S*'s identification with the entire act and evokes more independent arousal responses. Which sequence is preferable in future research depends on the purpose of the study. Our results indicate that sequence is a condition related to factor structure. For most research purposes, the random sequence may be preferred, since it yields factors more closely related to particular stimuli.

Of particular interest is the suggestion that individuals convicted of specific sexual offenses may react uniquely to one or more of the several arousal factors identified. For example, the greater arousal level of rapists to slides depicting heavy petting and intercourse and the low arousal level of the group convicted of indecent liberties with females to the fellatio slides are tentative findings which deserve further study. Upon further refinement, verbal report of sexual arousal to particular erotic stimuli might serve as a useful aid in the diagnosis and treatment of individuals with sexual maladjustments. Similarly, studies need to be done on the potential value of such response profiles in predicting subsequent sexual behavior. These results provide some initial suggestions.

The factor analyses indicate that a total of perhaps 5–10 slides could adequately represent the major sexual themes, eliminating the redundant slides (i.e., slides heavily loaded on the same factor). In this regard, the factor analyses presented can be useful in the development of subtests of erotic material and can aid the selection of the most diagnostic slides for each factor in future research.

REFERENCES

Brady, J. P., & Levitt, E. E. The relation of sexual preferences to sexual experience. *Psychological Record*, 1965, **15**, 377–384.

Cairns, R. B. *Psychological assumptions in sex censorship: An evaluative review of recent (1961–1968) research.* (Commission on Obscenity and Pornography, Tech. Rep., Vol. 1.) Washington, D. C.: United States Government Printing Office, 1971.

Gebhard, P. H., Gagnon, J. H., Pomeroy, W. B., & Christenson, C. V. *Sex offenders.* New York: Harper & Row, 1965.

Goldstein, M. J., Kant, H. S., Judd, L. L., Rice, C. J., & Green, R. *Exposure to pornography and sexual behavior in deviant and normal groups.* (Commission on Obscenity and Pornography, Tech. Rep., Vol. 7.) Washington, D. C.: United States Government Printing Office, 1970.

Johnson, W. I., Kupperstein, L., & Peters, J. *Sex offenders' experience with erotica.* (Commission on Obscenity and Pornography, Tech. Rep., Vol. 7.) Washington, D. C.: United States Government Pringint Office, 1970.

Report of the Commission on Obscenity and Pornography. Washington, D. C.: United States Government Printing Office, 1970.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Thorne, F. C., & Haupt, T. D. The objective measurement of sex attitudes and behavior. *Journal of Clinical Psychology*, 1966, **22**, 395–403.

Walker, C. E. *Erotic stimuli and the aggressive sexual offender.* (Commission on Obscenity and Pornography, Tech. Rep., Vol. 7.) Washington, D. C.: United States Government Printing Office, 1970.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

(Received September 11, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 512–520

RELATIONSHIP BETWEEN TASK DIFFICULTY AND THE CRITERION:

SHOULD WE MEASURE EARLY OR LATE?

EDWARD L. LEVINE AND JOSEPH WEITZ[1]

New York University

Task difficulty and the time at which we take our criterion measures seem to be interrelated as follows: if a task (or job) is "easy," the effect of an independent variable, in terms of optimal discriminability on a criterion measure, will be evidenced most clearly *early* in a task; if a task is "difficult," the effect of an independent variable will show up *later* in the task. The present research tests this formulation with groups as *S*s and corroborates it. In addition, a method of treating this hypothesis statistically is offered, along with a conjectural means of classifying tasks into easy or hard categories.

Weitz (1961) has suggested three major dimensions along which criteria may vary: (*a*) the time at which the criterion is measured, either early in the performance of a task (job) or late; (*b*) the level of difficulty of the criterion, easy or hard; and (*c*) the type of criterion, such as ratings, errors, etc.

This conceptualization takes on significance for two reasons. First, given the relatively rudimentary state of criterion measurement, it is evident that different psychological truths may emerge when any or all of the three criterion dimensions change. For example, a predictor may be related to a rating criterion but not to a more objective measure based on actual performance data. Second, an examination of the differences between results that arise when different proficiency measures are used, and their developmental trends, may shed some light on how the independent variable is working.

In attacking some of the problems that derive from this concept of criterion measurement, Weitz (1961, 1964, 1966) has shown both in his own research and by analysis of others' data that two of the criterion dimensions are interrelated as follows: If a task is "easy," maximal differentiation on a dependent variable will appear relatively *early* in the task; if a task is "hard," maximal differences will appear *late* in the task.

This relationship should hold, regardless of the independent variable employed in an investigation. That is, it should be clear that we are talking about the manner in which variables are interrelated, regardless of content. Furthermore, the hypothesis makes no statement about discriminability at other than the most "appropriate" times for measurement.

For example, consider the *difficult task* condition for a moment (but note that the argument presented below can be made, if properly modified of course, for easy tasks as well). An independent variable may have an effect early in a difficult task in terms of significance of difference on a criterion measure; and, when all trials are lumped together, there may be a significant mean difference as well. However, what is in question here is *relative discriminability*, or, the best point in time to take a criterion measure. We maintain that for difficult tasks, the best time to take a criterion measure is late in the task.

There are three conceivable alternatives if the dependent variable does indeed register a difference between two groups that vary on some independent variable, and if our hypothesis holds (again, for difficult tasks).

Case 1.

The most sizable discrimination between the groups occurs late, but there is a difference in the same direction early. This would suggest that an overall measure lumping all scores early and late would give an indication of a difference.

---

[1] Requests for reprints should be sent to Joseph Weitz, Department of Psychology, New York University, 21 Washington Place, New York, New York 10003.

Case 2.

The most sizable discrimination between the groups occurs late, but there is no difference between the groups early. This would suggest that late measurement is most likely to yield a significant difference. If all trial scores were lumped together, the difference might not attain significance.

Case 3.

The most sizable difference between the groups occurs late, but there is a trend in the early trials that is opposite to the significant difference obtained late. This would be more likely than Case 2 to cause a nonsignificant result when all trials are lumped. Also, if only early measures were taken, the researcher might form an erroneous impression of the operation of his independent variable.

In view of the fact that more powerful statistical tests (i.e., tests more likely to show significant results when the null hypothesis is indeed false) would result if measurements were taken late, and in view of the fact that it may not always be possible or desirable to gather criterion data at many points in time, it is clear that evidence in favor of the hypothesis outlined above concerning the best time to measure for difficult tasks has both practical and theoretical relevance. Both experimental and industrial psychologists could more efficiently plan their experiments and validity studies. In addition, theories may be more rigorously tested by usual statistical techniques if their power is increased by considering the point in time when a dependent variable is most likely to register a difference between groups that do differ on some independent variable.

Tasks may be easy or hard for a number of reasons, including complexity and ambiguity. Also, a given task may become easy or difficult due to the aptitude or suitability of the individual for the task. Thus, if an individual has high aptitude for a particular job, as measured by a valid test, then we might think of the task (job) for this individual as a relatively easy one. On the contrary, if an individual has low aptitude for the particular job, then the task might be considered more difficult for him. If the earlier formulation regarding task difficulty and dependent variable relationships is correct, this should mean that the effect of an independent variable will show up early for people with high aptitude and later for people with low aptitude. One example of this is that a job-relevant personality measure might be valid late for low-aptitude individuals and early for high-aptitude individuals.

Time of measurement, early or late, must be defined relative to the task that is undertaken. For experimental tasks used in a laboratory, early may mean the first few trials; for job performance, it may mean the first few months.

Several investigators have attempted to verify the hypothesis in its extension to personnel selection. Hakel (1968) demonstrated some support for the concept, and in the work by Dodd, Wollowick, and McNamara (1970) the hypothesis seems rather well supported. In both instances the validity of a personality measure appeared higher if aptitude was used as a moderator, and the personality measure related to performance at the appropriate points in time.

While this extension is interesting in and of itself, we are not testing it here. The task difficulty in the present experiment is defined independently of the suitability of the *groups* (groups are the *S*s in this study) for the task demands made upon them.

It should be noted that the time dimension of the criterion has been handled, but in a somewhat different context, by others (e.g., Fleishman, 1957; Fleishman & Fruchter, 1960; Fleishman & Hempel, 1954, 1955; Ghiselli & Haire, 1960). Their major orientation has been one of showing that relationships between predictors and performance measures change over time as a result of changes in the skill, or motivational, components of the job (task).

The present research proposes to test and extend Weitz' hypothesis. The criterion measures to be reported here are group performance measures as opposed to individual performance measures. This is, of course, a major departure from previous tests. In addition, the data were derived from a social psychological experiment with which Weitz had no connection (Levine, 1970). The principal investigator had no intention of testing the concept at the outset and, therefore, no expectations about outcomes. The analysis of the data and testing of the hypothesis was suggested after the study was completed.

The hypothesis to be tested is that the principle stated above, relating effectiveness of an independent variable and time and task difficulty, holds for groups as well as individuals. Discrimination between two sets of groups that differ on an independent variable will be greater relatively early (as opposed to late) when the groups are confronted with an easy task; but when groups are faced with a hard task, the differences will be greater late in the task.

METHOD

There are two independent variables to be considered, both of which are aspects of a group's control structure: the *amount* of control or influence possessed by group members over their group's decisions, and the *distribution* of control or influence in the group.

Four sets of three-man groups were set up as follows: Each member of the group was given role-playing instructions separately. These could be either "high-control" or "low-control" instructions.

The high-control instructions indicated that *S* should concentrate on maintaining a high rate of interaction and on *giving* suggestions. He should also try to avoid *asking* for opinions, information, and suggestions.

The low-control instructions indicated that *S* should concentrate on remaining relatively quiet, and when *S* did interact he would do so by *asking* for opinions, information, and suggestions. He was to *give* opinions, information, and suggestions only if asked to do so by the other group members. (Manipulation checks showed that *S*s could play the assigned role very well and perform the assigned tasks simultaneously. An outside observer was able to identify the high-control and low-control role players with over 80% accuracy, corrected for chance.)

> *Set 1.* 16 groups—high amount of control—balanced distribution.—All three *S*s received high control roles.
>
> *Set 2.* 16 groups—high amount of control—unbalanced distribution. Two of three *S*s received high-control roles, and the third received low-control instructions.
>
> *Set 3.* 16 groups—low amount of control—balanced distribution. All three *S*s received low-control roles.
>
> *Set 4.* 16 groups—low amount of control—unbalanced distribution. One of three *S*s received the high-control role, the other two had low-control roles.

In all sets, eight groups had easy tasks and eight had hard tasks. In comparing high versus low amount of control, Set 1 and Set 2 were compared to Set 3 and Set 4. In comparing balanced versus unbalanced distribution of control, Set 1 and Set 3 were compared to Set 2 and Set 4. These comparisons are based on an analysis of variance approach and may be considered orthogonal.

Task

The task is a modified version of one used by Hollander (1960). Three criteria were instrumental in its selection: (*a*) Does it hold *S*'s interest? (*b*) Does it yield definite and specifiable outcomes? (*c*) Is it readily manipulable in terms of ease and difficulty?

The task is centered around a 7 × 7 matrix (see Fig. 1) in which columns are designated by color names, rows by letters.

The elements are payoffs ranging from −12 to +15. Groups must decide upon a particular row for each of 15 trials. After the row has been determined, *E* declares the column for that trial—the column choices corresponding to a predetermined system. The object of the game is a maximization of payoffs to the group or, in other words, to figure out the system.

Instructions, taped, were as follows:

> Before you is a matrix of plus and minus numbers. The columns have color names and the rows have letter names. Your task is to choose one row for each of 15 trials to maximize your group's payoff, and to do this as quickly as possible.
>
> How is the payoff determined? It is the sum of plus and minus values you accumulate over 15 trials.
>
> How are these values determined? Here is an example. Trial 1. Your group chooses a row after a discussion and announces their choice to the experimenter. *Then* the experimenter announces the column color for that trial. The column color has been selected according to a predetermined system. For instance, the system may be consecutive columns. Thus, if "red" were the first column (Trial 1), then the color for Trial 2 would be "blue," for Trial 3, "yellow," and so on.
>
> Your best row choices under these conditions would be "F" for Trial 1, which yields a +15 with "RED", . . ., "D" for Trial 2 which yields a +10 with "blue", . . ., "G" for Trial 3 which yields a +10 with "yellow," . . ., and so on. The sum of your gains and losses is the payoff.
>
> Obviously, if you figure out the experimenter's system of choosing columns, your payoff will be a maximum, and you could complete all trials very quickly. The system and your row choices are therefore the main topics of discussion.
>
> Also, as part of your first trial you *must* discuss the following issues:
>
> 1. Which member of the group will announce the row choice to the experimenter.
> 2. Which member of the group will pose questions about the task, if you find that you don't understand what you have to do.
> 3. Whether supplies are necessary, for example, paper and pencil, and who will be responsible for these.
> 4. How the group's decisions are to be reached—by majority, unanimity, or some other method.

Row	Green	Red	Blue	Yellow	Brown	Orange	Black
A	−1	−12	+5	−1	−2	+15	−4
B	+10	−1	−2	−7	+4	−3	−1
C	−5	+5	−3	+3	−11	−1	+12
D	+5	−7	+10	−2	−5	+1	−2
E	−4	−1	−1	+1	+13	−10	+2
F	−6	+15	−5	−1	−3	−1	+1
G	−1	−1	−2	+10	+4	−2	−8

FIG. 1. The game matrix.

> A list of these issues will be given to you by the experimenter, and they are to be discussed during your first trial. Good luck!

The rationale for the form of the matrix and other procedural aspects is given by Hollander (1960, pp. 366–367).

The two output variables are time-to-completion of all trials, and the group's payoff.

Problem Difficulty

The means of making the problem easy or difficult obviously consists of alterations in the system of column choices for the trials. Hollander[2] suggests as a "simple" system, one with a "shift of two columns to the right on each trial, viewing black and green as adjacent in a circular arrangment." A more complex system might approach random selection of the column colors. After some preliminary piloting it was decided that each group would play two games with two separate systems. The easy systems were (e_1) three consecutive columns, skip one, and (e_2) two consecutive columns, skip two. The difficult (hard) systems were (h_1) 1, 3, 5, 7, 7, 5, 3, 1 columns forward from the color of the preceding trial, and (h_2) 1, 2, 3, 4, 5 columns backwards from the color of the preceding trial.

For each of the four problems, the initial color was randomly selected, and the rest followed from the system. Thus, the same two easy color sequences, and the same two difficult color sequences were used throughout. Maximum total payoff for the easy set of 30 trials is +360 and for the difficult set, +370; so any differences between the easy and difficult condition can hardly be due to differences in maximum payoff obtainable. Also, the order of presentation of the two problems was randomly determined for each group, with the constraint that the two possible orders for both easy and hard conditions be used an equal number of times. A group received either two easy or two hard problems.

Problem ease or difficulty should affect both output variables such that easy problems yield higher payoffs and take less time to complete than difficult ones. Furthermore, the "meat" of the group's discussion should be centered around the system and how to maximize payoffs within the system.

Subjects

The 64 groups were composed of three male or three female undergraduates enrolled in a basic psychology course. (Results are not reported separately for male and female groups since they did not differ on the overall group performance measure of current interest—group payoffs.) Participation in experiments was a requirement of the course. Subjects were randomly assigned to one of the four sets of groups and to the easy or hard problem conditions. Each experimental treatment in the design had equal numbers of male and female groups.

Procedure

Three subjects were taken into the lab and seated around a small table. They were told a little about the experiment and then given their role instructions individually in separate rooms with coaching when appropriate.

After a practice session, and more feedback in the separate rooms, they returned to the lab and the taped task instructions were played. They then completed 15 trials of Game 1. At this point they were taken into separate rooms once more and given feedback on their role playing. Game 2 was played, and 15 trials were again completed.

RESULTS

The task difficulty manipulation worked. Easy problems yielded higher payoffs and shorter completion times than did difficult problems (see Table 1). In the Results section

---

[2] E. P. Hollander, personal communication, June 4, 1969.

TABLE 1

CHECK ON THE EASY VERSUS DIFFICULT TASK MANIPULATION

Measure	Easy tasks	Difficult tasks	F
Time-to-completion[a]	31.43	44.46	9.05*
Payoff score[a]	291.44	178.72	40.87**

[a] Summed over both games. (A constant of 100 has been dded to payoff scores.)
* $p < .001$.
** $p < .0001$.

TABLE 2

OVERALL EFFECTS OF THE CONTROL STRUCTURE VARIABLES ON PAYOFF SCORES

Score	Amount of control		
	High	Low	F
$\bar{X}$ Payoff score[a]	260.19	209.97	8.11**
	Distribution of control		
	Balanced	Unbalanced	
$\bar{X}$ Payoff score[a]	255.59	214.56	5.41*

[a] Summed over both games. (A constant of 100 has been added to payoff scores.)
* $p < .03$.
** $p < .01$.

we are concerned only with the payoffs achieved by the groups for each trial since time scores reflected the length of the total game only, not the length of each trial.

In considering the data presented below, the reader might wish to know the effect on payoff scores over all trials of the control structure variables, amount of control, and distribution of control. Table 2 contains the data, which reveal that high amount of control groups scored significantly higher than low amount of control groups, and balanced distribution groups scored significantly higher than unbalanced distribution groups.

Graphical evidence bearing on the hypothesis of interest appears in Figures 2–5.

Figure 2, which depicts the performance (in blocks of three trials) of high- and low-control groups on easy tasks supports the hypothesis.

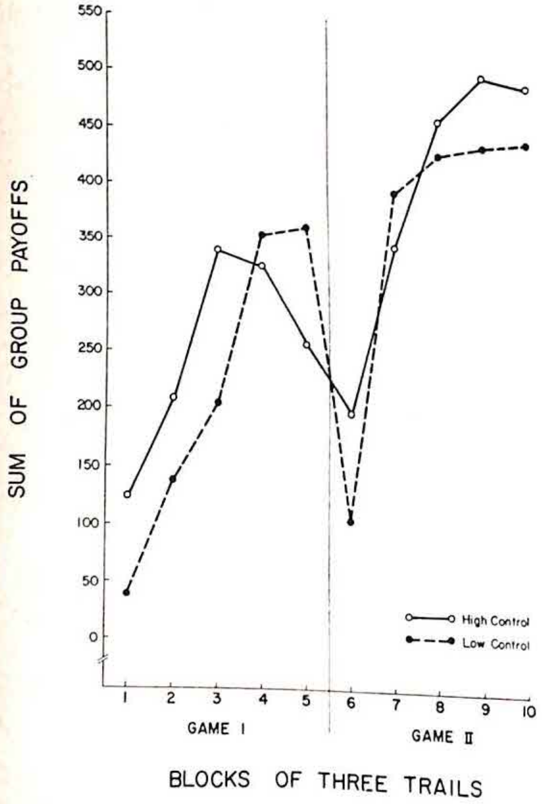


FIG. 2. Relationship between payoffs and trials for easy tasks (high–low control).

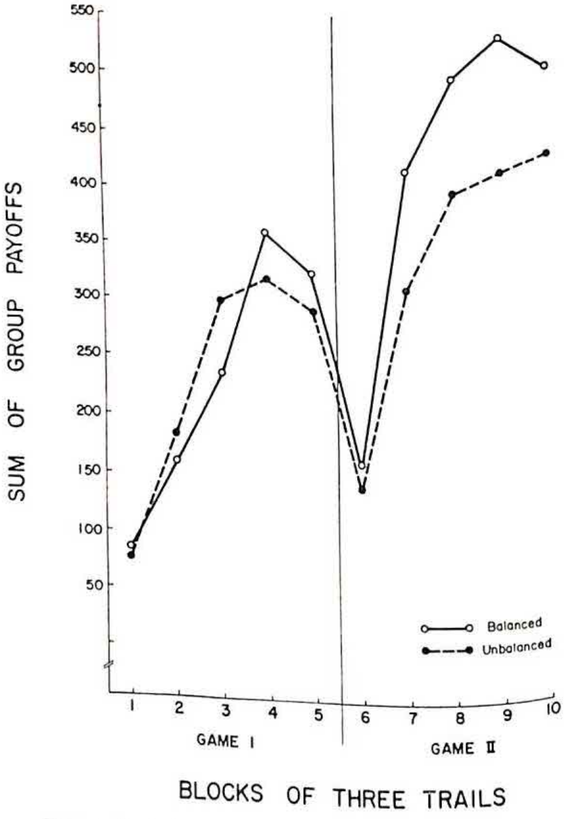


FIG. 3. Relationship between payoffs and trials for easy tasks (balanced-unbalanced).

The early trials of the first game offer the best discrimination between the high- and low-control groups. The evidence is not consistent, however, between first and second games since in the second game some difference appears relatively late. (It may be pointed out here again, as Dodd, Wollowick, & McNamara [1970] note, that the hypothesis does not predict an absence of discrimination late in an easy task, only that early discrimination is greater than late.)

When we compare the balanced distribution versus the unbalanced distribution sets of groups on the easy task (see Figure 3 again plotted in blocks of three trials), we find no consistent pattern in the first game.

However, in the second game, the *later* trials reveal a greater amount of differentiation, in which balanced groups are superior. This is contrary to the hypothesis.

Turning to the hard task condition, Figure 4 shows that the hypothesis as it related to more difficult tasks is supported.

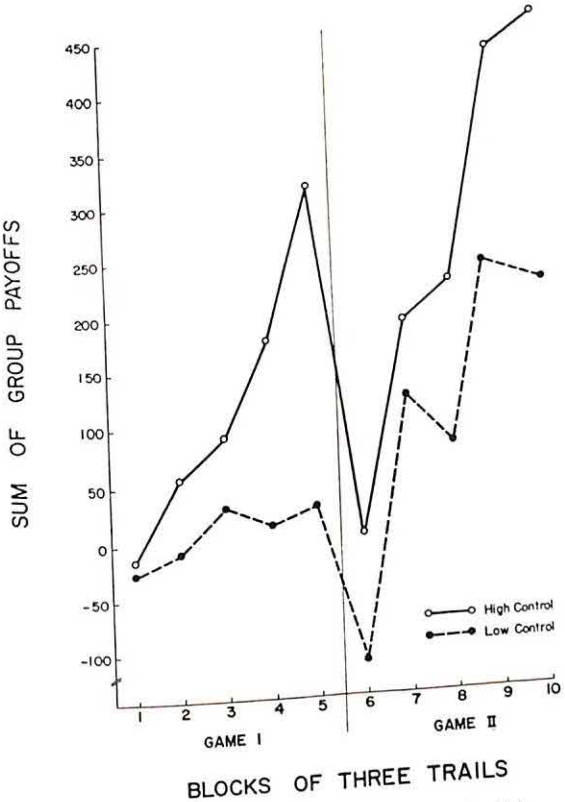


Fig. 4. Relationship between payoffs and trials for hard tasks (high-low control).

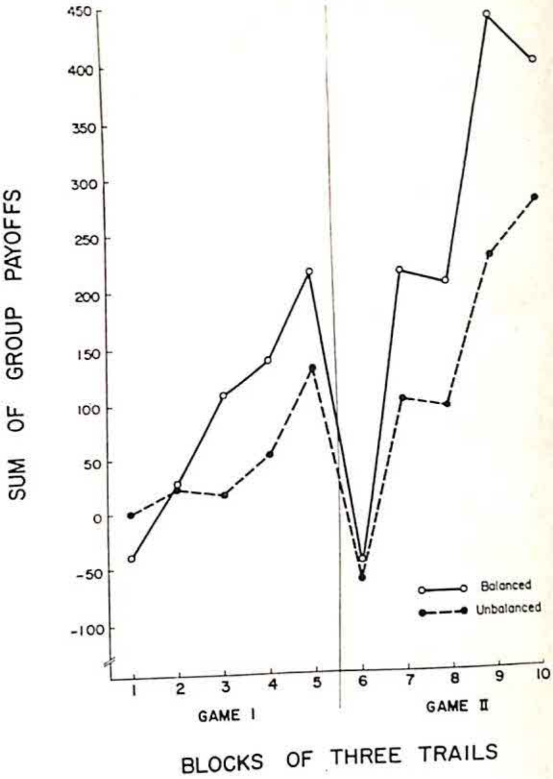


Fig. 5. Relationship between payoffs and trials for hard tasks (balanced-unbalanced).

The later trials in both the first game and the second game show the greatest differences between the two sets of groups with the high-control groups being clearly superior late in each of the sessions.

Figure 5 also demonstrates that the block trial data for the later trials for Games 1 and 2 (hard task—balanced versus unbalanced distribution) are more discriminating than the early trials. The balanced groups show consistently better performance late in each game.

Statistical Treatment of the Data

In an attempt to develop a statistical technique to handle the results of these kinds of problems, it was decided to compute a t ratio for each trial (1–15, for both games).

These t ratios can then be compared in two ways:

1. The mean t ratios for Trials 1–8 for Games 1 and 2 (16 scores) can be tested against the mean t ratio for Trials 9–15 for Games 1 and 2 (14 scores). It should be noted that this divi-

sion, first eight versus second seven, is not necessarily the best one possible, especially for the hard condition. It was used merely to split the games approximately into first half versus second half. The significance tests might therefore be considered somewhat conservative.

2. Another way of comparing the *t* ratios also stems from the hypothesis. Since the two games in each condition differ only in the systems, both games (30 trials altogether) might be considered one task. The first half of the first game would then be early, and the rest of the trials would be late. Therefore, the first eight *t* ratios of Game 1 may be contrasted with those of the sum of the last seven trials of Game 1 and all 15 of Game 2.

In view of the high degree of similarity between Games 1 and 2, the comparison between the first 8 and last 22 *t* ratios seems most apposite as a test of the theory, although both kinds of comparisons are presented and discussed to assess the possibility that the two produce different results. This use of *t* ratios may be seen as a rather novel way of handling the question of *relative discriminability*, that is, significant mean differences between sets of *t* ratios indicates that one set shows greater differentiation between the experimental groups of interest. (As required by the *t* test, both variance estimates in each comparison were tested. All pairs were homogeneous.)

The results of these comparisons offer some support for the hypothesis.

Table 3 contains the results of these comparisons for the easy task condition.

The data reveal that for the *primary* comparison (first 8 versus last 22 *t* ratios), the high-control groups are significantly more successful, compared to low-control groups in the *early* going. Groups with a balanced control structure, however, are significantly more successful than those with unbalanced structures *later* in the tasks. In fact, if we confined our criterion measurement to the first eight

TABLE 3

COMPARISONS OF THE MEANS OF EARLY SETS OF *t* RATIOS AND THE MEANS OF LATE SETS OF *t* RATIOS FOR EASY TASKS

Comparison	Early	Late	t^a
I. High versus low amount of control comparison			
Primary comparison			
First 8 trials versus last 22 trials			
$\bar{X}$	$t_8 = .84$	$t_{22} = .04$	2.17***
SD	$t_8 = .94$	$t_{22} = .96$	
Secondary comparison			
First half versus second half of both games			
$\bar{X}$	$t_{16} = .45$	$t_{14} = .02$	1.24*
SD	$t_{16} = .94$	$t_{14} = .96$	
II. Balanced versus unbalanced distribution			
Primary comparison			
First 8 trials versus last 22 trials			
$\bar{X}$	$t_8 = -.22$	$t_{22} = .63$	−2.47****
SD	$t_8 = .76$	$t_{22} = .86$	
Secondary comparison			
First half versus second half of both games			
$\bar{X}$	$t_{16} = .17$	$t_{14} = .66$	−1.52**
SD	$t_{16} = .82$	$t_{14} = .95$	

Note.— Abbreviation t = trial.

[a] The minus sign indicates that the direction of the difference is the opposite of that predicted.

* $p < .25$ (two-tailed test).

** $p < .15$ (two-tailed test).

*** $p < .05$ (two-tailed test).

**** $p < .02$ (two-tailed test).

TABLE 4

Comparisons of the Means of Early Sets of *t* Ratios and the Means of Late Sets of *t* Ratios for Hard Tasks

Comparison	Early	Late	*t*
I. High versus low amount of control			
Primary comparison			
First 8 trials versus last 22 trials			
$\bar{X}$	$t_8 = .359$	$t_{22} = 1.496$	2.872***
SD	$t_8 = .934$	$t_{22} = .968$	
Secondary comparison			
First half versus second half of both games			
$\bar{X}$	$t_{16} = .668$	$t_{14} = 1.793$	3.3275***
SD	$t_{16} = .965$	$t_{14} = .876$	
II. Balanced versus unbalanced distribution			
Primary comparison			
First 8 trials versus last 22 trials			
$\bar{X}$	$t_8 = .126$	$t_{22} = .948$	2.001**
SD	$t_8 = .828$	$t_{22} = 1.047$	
Secondary comparison			
First half versus second half of both games			
$\bar{X}$	$t_{16} = .487$	$t_{14} = 1.005$	1.372*
SD	$t_{16} = 1.207$	$t_{14} = .780$	

Note.—Abbreviation *t* = trial.
* $p < .20$ (two-tailed test).
** $p < .06$ (two-tailed test).
*** $p < .01$ (two-tailed test).

trials, we would find, contrary to the overall results, that the unbalanced structure groups were superior to the balanced structure groups.

The data of the first-half versus second-half comparison (the *secondary* comparison) yields a similar picture to the primary results with one supportive and one contradictory result for the amount of control and distribution of control variables, respectively, although the results here do not approach conventional significance levels.

Thus, the hypothesis is supported in one instance and contraverted in another. It may be possible, although on an *ex post facto* basis only, to resolve this negative finding in favor of the hypothesis based on the suitability of the various group structures for the tasks they faced. The matter is explored further in the discussion.

Table 4 contains the data for the results of the *t* ratio comparisons for the hard task condition. Here the primary comparisons (first 8 versus last 22) are both supportive of the hypothesis. Thus, high amount of control and balanced distribution of control groups are relatively more successful late in the task than their counterparts. The secondary comparisons show similar results, though in one of the two cases (balanced versus unbalanced) the *t* ratio difference does not approach the conventional alpha level.

Discussion

Both the graphic and statistical data offer support for the hypothesis. It seems, therefore, that within the limits of our experiment early measurement is better suited than late measurement to unearth differences between groups, when tasks are easy; when tasks are hard we should measure performance later in the task. This was so for all comparisons but one, the balanced versus unbalanced distribution condition for the easy task. What of this discrepancy?

In answer to this problem, recall that at the outset, we stated that an examination of the variations in performance at various stages in the task might shed some light on how the independent variables operate. This is borne out to a degree by the contrary findings.

The discrepant results of the balanced versus unbalanced groups on easy tasks stem primarily from the performance of one set of groups, the low-control, balanced distribution condition. Since everyone in these groups had a low-control role, there was little movement toward the goals of quick completion of trials and maximum payoffs. In essence any task would be difficult for this type of group, and their scores served to depress the scores of the balanced set of groups as a whole in the early going. In fact, the significant results of Table 3, which contravert Weitz's hypothesis, indicate that this comparison behaved in a manner to be expected of hard tasks, that is, maximal differentiation later in the task. This is an admittedly *ex post facto* explanation, but it should serve to illustrate that a task may indeed be made more or less difficult as a function of individual or group suitability for the task. In terms of Weitz's hypothesis, this explanation of negative results would most appropriately be considered under the heading of the postulate or extension to the hypothesis, although, again, we were not explicitly testing it in the present study.

In addition, our understanding of the independent variables is furthered by an examination of the graphic evidence (see Fig. 2–5). There seems to be an interaction between the independent variables (amount of control and distribution of control) and the ease or difficulty of the task; differences in control structure become more important when the task is difficult.

Turning now to more general considerations, we find it apparent that the definition of easy and hard tasks is rather ad hoc. This would complicate the life of an investigator who wished to know whether to measure early or late. The programmatic work of Fleishman (summarized in Fleishman, 1966) and the study by Ghiselli and Haire (1960) might have some bearing on this by offering some means of classifying tasks based on the qualitative and quantitative shifts in ability and motivational components at various stages of task performance. Easy tasks might undergo fewer shifts than difficult tasks. Fleishman in particular has shown that late stage proficiency is harder to predict than early proficiency and has sought to identify variables especially predictive of late proficiency. The present study perhaps contributes in showing that task difficulty may be an important moderator in such predictions.

At any rate, the conceptualization of criterion measurement described here seems a heuristic one and might lead to a better understanding of what the independent variable is actually doing. It may also serve as a starting point for a concerted attack on that most important problem—the criterion.

REFERENCES

DODD, W. E., WOLLOWICK, H. B., & MCNAMARA, W. J. Task difficulty as a moderator of long-range prediction. *Journal of Applied Psychology*, 1970, **54**, 265–270.

FLEISHMAN, E. A. A comparative study of aptitude patterns in unskilled and skilled psychomotor performances. *Journal of Applied Psychology*, 1957, **41**, 263–272.

FLEISHMAN, E. A. Individual differences and motor learning. In R. M. Gagne (Ed.), *Learning and individual differences*. Columbus: Charles E. Merrill, 1966.

FLEISHMAN, E. A., & FRUCHTER, B. Factor structure and predictability of successive stages of learning Morse code. *Journal of Applied Psychology*, 1960, **44**, 96–101.

FLEISHMAN, E. A., & HEMPEL, W. E. Changes in factor structure of a complex psychomotor test as a function of practice. *Psychometrika*, 1954, **18**, 239–252.

FLEISHMAN, E. A., & HEMPEL, W. E. The relation between abilities and improvement with practice in a visual discrimination reaction task. *Journal of Experimental Psychology*, 1955, **49**, 301–312.

GHISELLI, E. E., & HAIRE, M. The validity of selection tests in the light of the dynamic character of criteria. *Personnel Psychology*, 1960, **13**, 225–232.

HAKEL, M. D. Task difficulty and personality test validity. *Psychological Reports*, 1968, **22**, 502–511.

HOLLANDER, E. P. Competence and conformity in the acceptance of influence. *Journal of Abnormal and Social Psychology*, 1960, **61**, 365–369.

LEVINE, E. L. The effects of variations in control structure upon small-group performance and satisfaction: A laboratory study. Unpublished doctoral dissertation, New York University, 1970.

WEITZ, J. Criteria for criteria. *American Psychologist*, 1961, **16**, 228–231.

WEITZ, J. The use of criterional measures. *Psychological Reports*, 1964, **14**, 803–817.

WEITZ, J. Criteria and transfer of training. *Psychological Reports*, 1966, **19**, 195–210.

(Received July 20, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 521–525

RELATIVE INFLUENCE OF JOB AND SUPERVISION ON SHARED WORKER ATTITUDES[1]

DOUGLAS B. CAMPBELL[2]

Survey Research Center, University of Michigan

The question addressed is whether similarities of job attitudes among workers are more closely associated with the nature of the job itself or with the nature of the supervision received. Comparisons of attitude similarity between pairs of workers were made for four classes: workers having the same job as one another, the same supervisor, both, or neither. The *S*s were 311 workers from a wide variety of occupations in the southern Michigan area. Comparisons were made between class means to assess the relative association of job and supervision with attitude similarity between workers. Overall results showed shared supervision to be significantly associated ($p < .05$) with attitude similarity, while shared job was not significantly associated. It was suggested that findings could be understood by linking Maslow's hierarchy of needs with job and supervision.

Many studies have explored the relative relevance and importance to worker attitudes of various factors in the work environment. Herzberg, Mausner, Peterson, & Capwell (1957) and Vroom (1964), among others, have reviewed and summarized much of this literature. Herzberg et al. present data showing that the factor mentioned most often in 150 studies of worker attitudes and satisfactions was "intrinsic aspects of the job," followed by "supervision," and then "working conditions," thus establishing "supervision" as a factor independent of other "job" factors as well as suggesting the relative importance of the factors. Further data in Herzberg et al. show the same rank order of these three factors when their average importance is compiled from 16 studies involving over 11,000 workers. These tables may be interpreted as giving a slight edge to the job itself, while other sources (e.g., Ash, 1954; Ott, 1965; Putnam, 1930) suggest the greater influence of supervision. Still other studies report one view but are open to other interpretations than those made by the original investigator. For example, Morse (1953) found a relationship between attitude toward supervision and general morale. To test whether it was a halo effect or supervisors actually were affecting overall morale, she compared work groups having the highest average morale index scores with those having the lowest scores, for all measures of attitude toward supervision. Only two of the nine comparisons were significant ($p < .05$, two-tailed t tests), and she concluded it was merely a halo effect. Inspection of her tables, however, shows that *all* differences were in the predicted direction. Furthermore, her hypothesis dictated a one-tailed test, in which Case 3 of the nine tests were significant at the .05 level and six of the nine were significant at the .10 level. The supervisor may have had a more important effect on morale than Morse allowed. In general, then, no conclusive findings have emerged indicating the relative influence of various work factors, especially the factor of job itself and that of supervision, on a worker's attitudes.

Demographic variables complicate matters further. Herzberg et al. (1957) report that the relative influence of the job itself as opposed to supervision on worker attitudes is confounded by such variables as education, occupation level, income, and sex.

In an attempt to clarify past findings, this study examined the relative importance to worker attitudes of the job itself versus supervision. The study employed a somewhat different methodology from those used in the past. Many past studies have been concerned primarily with relationships between various work factors and "high satisfaction" or "good attitudes." This study assessed the extent to which the job itself and the supervision a worker receives are associated with workers' holding similar attitudes and perceptions as one another, be those attitudes and perceptions good or bad. This study also differed from most previous ones in the widely diverse occupations included in the population studied.

---

[1] Data for this report were gathered as part of a study supported by a grant from the United States Department of Labor (L-68-84).

The author is indebted to Robert P. Quinn and Stanley E. Seashore for helpful comments made in preparation of this manuscript.

[2] Requests for reprints should be sent to Douglas B. Campbell, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

METHOD

Subjects

Data were obtained from interviews with 311 workers in the southern Michigan area. These *S*s were selected to cover a very wide range of occupations and not as a representative sample. As a result, *S*s were somewhat overrepresentative of males, white-color workers, older workers, the more highly educated, and those in higher status occupations.

Part of the population interviewed was composed of preselected pairs of workers, that is, workers having the same job with the same employer, or having the same supervisor, or both. If *S* had been designated prior to his interview as a pair member, the interviewer verified this at the close of the interview by asking *S* if his assigned pair mate did indeed have the same job with the same employer and have the same supervisor. Additional pairs were generated by asking each unpaired *S* if he could name someone fulfilling the pair qualifications listed above. In this way, four classes of *S* pairs were generated: (*a*) a class in which the pair members had the same job and the same supervisor as one another (SJ/SS); (*b*) a class in which the pair members had the same job as one another but different supervisors (SJ/DS); (*c*) a class in which pair members had different jobs but the same supervisor (DJ/SS); and (*d*) a class comprised of random pairings of the rest of the *S*s, giving those pairs different jobs and different supervisors (DJ/DS). Since the original number of *S*s was 311, one *S*, chosen at random, was deleted from the analysis.

Procedure

Deriving the indexes. The *S*s were interviewed by professional interviewers employed by the Survey Research Center of the University of Michigan. Among the data obtained from these interviews were the 13 indexes of worker attitudes and satisfaction used in the subsequent analysis. Six of the indexes came from a set of items designed to assess job satisfaction. The *S* was asked how true each of these items was of his job (e.g., "the work is interesting," "I have enough time to get the job done"). Four alternative answers could be chosen: "very true," "somewhat true," "not too true," or "not at all true." The items were then combined on the basis of a factor analysis that yielded the six satisfaction indexes. The *S* was also asked how true each of several phrases was of his immediate supervisor (e.g., "is friendly and approachable," "checks frequently on the work of those he supervises"). These items also were factor analyzed and yielded four supervision indexes.

Three other indexes were included in the analysis. One was an index of job challenge, derived from a series of questions concerning how much *S*'s job required or allowed him to do a lot of planning ahead, have a high level of skill, be creative, do a variety of different things, and have to keep learning new things. Another index was composed of "content-free" job satisfaction questions (e.g., would *S* recommend his job to a friend?, would *S* take the same job if given a chance to do things over again?). The final index was the job tension index constructed by Kahn, Wolfe, Quinn, Snoek, and Rosenthal (1964).

Controlling for item reference. In order to control for the extent to which the items in each index were more job referent than supervisor referent, or vice versa, an attempt was made to categorize each index on the basis of the predominant reference of its items. Three colleagues who were not familiar with the aims of the study were asked to rate each item on a 7-point scale according to the degree to which they felt a "typical respondent" would answer that item with reference to his job itself, his supervisor, or some combination of the two. One end of the scale was labeled "job referent only," the other end, "supervisor referent only," and the midpoint was labeled "equal reference." Each index was then assigned a score ranging from 1.00 to 7.00, based upon the mean of the ratings of all items in that index by the three judges. Interrater agreement, computed according to Guilford (1954), was .67.

The indexes were then divided into two sets, based upon the judges' ratings. Those indexes having a score of less than 4.00 were placed in a set called job-referent half, while those with a score of greater than 4.00 were placed in the supervision-referent half Indexes receiving predominantly supervision-referent ratings (scores greater than 4.00) were the four indexes from the supervision factor analysis and one index from the job satisfaction factor analysis which measured satisfaction with supervision. The other indexes—the five remaining job satisfaction indexes, the content-free job satisfaction index, and the job tension index—received predominantly job-referent ratings (scores ess than 4.00). The job challenge index received a rating of exactly 4.00 and was not used in analyses of the two sets.

Computing the pair scores. The scores for each *S* on each index were derived by taking the mean score of all items in the index. After scores on each index for each *S* had been obtained, the absolute difference between the scores of each *pair* of *S*s was computed for each index. This absolute difference represented the extent of attitudinal dissimilarity between members of a given pair. The absolute differences between paired *S*s were the basic data that were used in the analysis.

The mean of the absolute difference scores (M_{AD}) for the pairs comprising a given class was then computed, providing the average absolute discrepancy between workers characterized as having either (*a*) the same job and the same supervisor, (*b*) the same job but different supervisors, (*c*) different jobs but the same supervisor, or (*d*) different jobs and different supervisors (see Table 1).

TABLE 1

MEAN ABSOLUTE DIFFERENCE (M_{AD}) SCORES FOR EACH INDEX BY CLASS

Index	M_{AD} scores			
	Class 1 (SJ/SS)[a]	Class 2 (SJ/DS)[b]	Class 3 (DJ/SS)[c]	Class 4 (DJ/DS)[d]
Job-referent indexes				
Job satisfaction 1	.48	.60	.45	.53
Job satisfaction 2	.38	.56	.53	.52
Job satisfaction 3	.55	.61	.80	.60
Job satisfaction 4	.53	.77	.53	.83
Job satisfaction 5	.49	.75	.25	.60
Content-free job satisfaction	.69	.71	.94	.91
Job tension	.58	.84	.53	.66
Equal-reference Index				
Job challenge	.32	.56	.53	.74
Supervision-referent indexes				
Supervision 1	.43	1.05	.50	.66
Supervision 2	.68	.28	.58	.85
Supervision 3	.66	1.25	.63	.57
Supervision 4	.55	.32	.66	.71
Job satisfaction 6	.66	.52	.49	.74
M_{AD}—Total array	.54	.68	.57	.69
M_{AD}—Job-referent half	.53	.69	.58	.66
M_{AD}—Supervision-referent half	.60	.68	.57	.71

Note.—Abbreviations: SJ = same job, SS = same supervisor, DJ = different job, DS = different supervisor.

[a] $n = 42$ pairs.

[b] $n = 9$ pairs.

[c] $n = 8$ pairs.

[d] $n = 96$ pairs.

Assessing job versus supervisor influence. To assess the actual relative importance of the job as opposed to supervision, the following between-class comparisons between M_{AD}s were made for the total array of indexes and for the job-referent and supervisor-referent sets as rated by the judges:

Comparisons assessing job influence were: (a) M_{AD}(DJ/SS) versus M_{AD}(SJ/SS) and (b) M_{AD}(DJ/DS) versus M_{AD}(SJ/DS).

Comparisons assessing supervisor influence were: (c) M_{AD}(SJ/DS) versus M_{AD}(SJ/SS) and (d) M_{AD}(DJ/DS) versus M_{AD}(DJ/SS).

As can be seen in the job influence comparisons, in each instance the supervisor factor remained constant between classes (either SS or DS in each class) while the job factor changed. The opposite was true in the supervisor influence comparisons; the job factor remained constant in each comparison while the supervisor factor varied. In comparisons a and b, then, to the extent that differences between M_{AD}s were significant, an association was indicated between workers' jobs and their attitudes. Likewise, in comparisons c and d significant differences between means indicated the association of workers' supervisors with those workers' attitudes These comparisons were made for each of the 13 indexes separately (for purposes of sign tests), for the indexes pooled overall, and for the two sets of indexes, pooled into job- and supervisor-referent halves as described above.

RESULTS

For each of the 13 indexes, t tests were run for all four types of comparisons such that a positive t value always indicated results in the predicted direction. A sign test (Hays, 1963) was performed on the t values for the comparisons designed to assess supervisor influence and was significant at the .02 level. A second sign test was performed for the job influence comparisons. The proportion of these comparisons that turned out in the predicted direction, however, was not significantly greater than one would expect by chance; in fact, it was slightly less than .50.

For each of the four classes represented by the columns in Table 1, the M_{AD} was computed for the entire array of indexes ($n=13$), the job-referent half ($n=7$) and the supervision-referent half ($n=5$). These means are shown at the bottom of Table 1.

The t-test comparisons for job and supervision influence were then made among the four classes. The results are shown in Table 2.

TABLE 2

CLASS COMPARISONS ON POOLED INDEXES (t VALUES)

Pooled indexes	Job influence comparisons		Supervision influence comparisons	
	Class 3 (DJ/SS) versus Class 1 (SJ/SS)	Class 4 (DJ/DS) versus Class 2 (SJ/DS)	Class 2 (SJ/DS) versus Class 1 (SJ/SS)	Class 4 (DJ/DS) versus Class 3 (DJ/SS)
Total array	.57	.09	1.74*	2.01*
Job-referent half	.51	−.40	3.08**	.86
Supervision-referent half	−.41	.11	.43	2.33*

Note.—Abbreviations: SJ = same job, SS = same supervisor, DJ = different job, DS = different supervisor.
* $p < .05$ (one-tailed test).
** $p < .01$ (one-tailed test).

The first row of the table shows that, overall, having supervision in common was much more highly associated with shared attitudes between workers than was having the same job as one another. Both comparisons assessing the influence of supervision were significant beyond the .05 level, while neither comparison assessing job influence was significant. Results in the last two rows show that, even with the predominant reference of the index items controlled by dividing the indexes into job- and supervision-referent halves, no comparison assessing job influence was significant. Regarding supervisor influence, one of the two comparisons for both the job-referent half and the supervision-referent half was significant at least beyond the .05 level.

DISCUSSION

How can one account for this consistent association of common supervisor with attitude similarity, coupled with a consistent *lack* of association of common job and attitude similarity? It is possible that the distinction made between factors that are "job influenced" as opposed to those that are "supervisor influenced" is mainly artificial. Reconsider the extent to which a supervisor may indeed be expected to influence various job factors. Myers (1964) points out the importance of the supervisor in satisfying both motivation and maintenance needs. Indeed, many of Herzberg's satisfiers and dissatisfiers (e.g., recognition, responsibility, advancement, interpersonal relations with the supervisor) can be viewed as heavily, if not completely, mediated by the supervisor. And, too, these factors mediated by the supervisor have been shown to be more important for both satisfaction and dissatisfaction than other job dimensions not so likely to be mediated by the supervisor (Dunnette, Campbell, & Hakel, 1967). Furthermore, these findings held up across several diverse occupational categories, such as are found in the current study. Friedlander's (1963) factors of job satisfaction also contain many items that seem under partial, if not entire, control of the supervisor.

But a more fruitful explanation seems to be one related to the well known hierarchy of needs proposed by Maslow (1954). Consider the following argument: Lower order needs are met mainly through the job itself. Highest order needs (self-actualization and, to some extent, esteem) also are met mainly through the actual job. But the intermediate needs (belongingness, love, and such esteem factors as the desires for prestige, recognition, attention, and appreciation) can be met to a great extent only by people, and, in the work environment, especially the supervisor. Furthermore, the supervisor probably affects both the amount of the need and the degree to which the need is satisfied, which combine to produce satisfaction, according to Morse (1953) among others. If one grants the above argument and then examines the findings of this study, the conclusion may be that, at this point in time, workers of the kinds studied are generally at the stage of meeting middle-to upper-level needs such that job characteristics are not so influential as they once were.

The job environment of workers at all levels has been improved, especially at the higher levels that are prevalent in this sample. Haire, Ghiselli, and Porter (1966) found greater satisfaction and fulfillment of needs in higher level than in lower level managers in the United States. Blai (1964) found Maslow's higher level needs selected as important by workers with increasing frequency from low in trades-manual and service occupations to high in professional occupations. Lower level needs were selected as important more frequently by *S*s in low socioeconomic status groups. Following this reasoning, one might expect that in some other countries less economically advanced than the United States, need satisfaction would be lagging comparatively, and this should be expected to result in a different order of salience than is found in a more advanced economy. This seems to be confirmed by Lahiri (1965), who found salary and security to be judged as the two most important job factors by both government and nongovernment clerical white-collar workers in India. "A good and sympathetic supervisor" was ranked consistently low in importance. Lahiri & Choudhuri (1966) reported similar results for technical and nontechnical employees of an industrial organization in Calcutta. In addition, an international study by Haire et al. (1966) found Indian managers reporting the highest degree of fulfillment of security needs but the lowest fulfillment among managers in all 14 countries of the sample for esteem and autonomy needs and the second lowest for fulfillment of self-actualization needs.

Maslow (1954) proposed that "needs cease to play an active determining or organizing role as soon as they are gratified [p. 105]." If the above argument has any validity, it is little wonder that the job itself, if it mainly meets already gratified needs, may show little association with worker attitudes. It is likewise easy to see why the supervisor, who plays a key role in meeting the needs speculated as being most salient at this time, is strongly associated with those same attitudes and perceptions. One might suggest that work environments will continue to improve, meeting higher and higher level needs, and if the above argument indeed holds, then as time passes we shall witness the reemergence of the job characteristics as the critical factors in worker attitudes through their relationship to the highest order needs.

REFERENCES

ASH, P. The SRA employee inventory—A statistical analysis. *Personnel Psychology*, 1954, **7**, 337–364.

BLAI, B., JR. An occupational study of job and need satisfaction. *Psychological Reports*, 1964, **14**, 81–82.

DUNNETTE, M. D., CAMPBELL, J. P., & HAKEL, M. D. Factors contributing to job satisfaction and job dissatisfaction in six occupational groups. *Organization Behavior and Human Performance*, 1967, **2**, 143–174.

FRIEDLANDER, F. Underlying sources of job satisfaction. *Journal of Applied Psychology*, 1963, **47**, 246–250.

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

HAIRE, M., GHISELLI, E. E., & PORTER, L. W. *Managerial thinking: An international study*. New York: Wiley, 1966.

HAYS, W L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

HERZBERG, F., MAUSNER, B., PETERSON, R. O., & CAPWELL, D. F. *Job attitudes: Review of research and opinion*. Pittsburgh: Psychological Services of Pittsburgh, 1957.

KAHN, R. L., WOLFE, D. M., QUINN, R. P., SNOEK, J. P., & ROSENTHAL, R. A. *Organizational stress: Studies in role conflict and ambiguity*. New York: Wiley, 1964.

LAHIRI, D. K. Perceived importance of job factors by government and nongovernment employees. *Indian Journal of Psychology*, 1965, **40**, 37–47.

LAHIRI, D. K., & CHOUDHURI, P. K. Perceived importance of job factors by technical and nontechnical employees. *Personnel Psychology*, 1966, **19**, 287–296.

MASLOW, A. H. *Motivation and personality*. New York: Harper & Row, 1954.

MORSE, N. C. *Satisfactions in the white-collar job*. Ann Arbor: Survey Research Center, Institute for Social Research, University of Michigan, 1953.

MYERS, M. S. Who are your motivated workers? *Harvard Business Review*, 1964, **42**(1), 73–88.

OTT, D. C. The generality of Herzberg's two-factor theory of motivation. Unpublished doctoral thesis, Ohio State University, 1965.

PUTNAM, M. L. Improving employee relations. *Personnel Journal*, 1930, **8**, 314–325.

VROOM, V. H. *Work and motivation*. New York: Wiley, 1964.

(Received August 10, 1970)

ournal of Applied Psychology
971, Vol. 55, No. 6, 526–530

PERFORMANCE AND KNOWLEDGE OF RESULTS AS DETERMINANTS OF GOAL SETTING[1]

L. L. CUMMINGS,[2] DONALD P. SCHWAB, AND MARC ROSEN[3]

Graduate School of Business, and Industrial Relations Research Institute, University of Wisconsin

Previous performance on a simple addition task was found to exert a significant positive impact on the goal levels set by 80 *S*s. When performance effects were accounted for, knowledge of results (KR) also influenced goals significantly. When the effects of four forms of KR were compared, correct KR increased goal level significantly above that generated by no KR, while incomplete KR increased goal level insignificantly, and erroneous low KR decreased performance below the level associated with no KR. Implications of the findings for administrative practice and goal-setting theory were discussed.

The question of the motivational determinants of work performance is now one of the primary concerns in the literature of applied psychology (e.g., Porter & Lawler, 1968; Schwab & Cummings, 1970; Vroom, 1964). Several theoretical orientations have been taken on this issue. The two-factor theory (Cummings & El Salmi, 1968; Herzberg, 1966; Herzberg, Mausner, & Snyderman, 1959), for example, emphasizes the role of job satisfaction in determining performance levels. In expectancy theory (Porter & Lawler, 1968; Vroom, 1964) a crucial role is played by perceptions of the relationship between performance and the receipt of valued outcomes. Equity theory (Adams, 1965; Pritchard, 1969) emphasizes social comparisons and the impact of feelings of dissonance on performance.

Recently Locke and his associates (Locke, 1968a, 1969) have emphasized the importance of conscious *goal levels* as determinants of employee performance (Locke & Bryan, 1969a, 1969b; Locke, Cartledge, & Koppel, 1968). In summarizing his work on this issue Locke (1968b, p. 157) has argued that: (*a*) hard goals produce a higher level of performance (output) than easy goals, (*b*) specific hard goals produce a higher level of output than a goal of "do your best," and (*c*) behavioral intentions regulate choice behavior.

Locke (1968b) has also argued that goals and intentions mediate the effects of extrinsic rewards such as monetary incentives, knowledge of results, and time limits on performance. In general it is argued "that external incentives such as instructions or feedback do not affect behavior independent of their effects on goal setting or intentions [Locke, Cartledge, & Knerr, 1970, p. 152]."

If goals determine performance and mediate the effects of extrinsic factors, it is nevertheless possible for various types of external incentives to influence the level of goals set. Although Locke et al. (1970) have suggested that their "model does not delineate the factors which influence the establishment and acceptance of (end) goals and general values . . . in the first place [p. 154]," it is interesting to note that at least one study (Locke & Bryan, 1968) suggests that knowledge of results (KR) may influence the level of goals set. In that study two groups (KR and no KR) performed on a computation task for 16 5-minute trials. At the end of the eighth, before the ninth, and at the end of the sixteenth trials, *S*s were asked to specify goals on a 10-item goal check list. The authors report that the goals specified on the third questionnaire were significantly related to KR, higher goals being associated with the group that received KR.

Taken at face value these results suggest that correct KR results in higher goals. There are several limitations of the study, however, which reduce the confidence one can place in the findings and thus suggest the need for additional research on this issue. First, goal items were specified in potentially ambiguous terms (e.g., "working along at a reasonably

---

[1] The authors would like to acknowledge the assistance of Leonard Berkowitz, Lee Dyer, and William Devitt in conducting this study.

[2] Requests for reprints should be sent to L. L. Cummings, Graduate School of Business, University of Wisconsin, 1155 Observatory Drive, Madison, Wisconsin 53706.

[3] Now at Northrop Corporation, Los Angeles, California.

fast pace but not the fastest possible"). Second, scores of *S*s who responded to 3 of the 10 items were dichotomized and correlated with KR. This procedure resulted in the elimination of an unknown number of *S*s who checked one of the other seven goal items. Third, the relation between KR and goals was described for the three goal statements, but statistical results are only presented for the last one following Trial 16.

Finally, inadequate account was taken of the potential relationship between previous performance, KR, and goals. Specifically, previous performance may be instrumental in the establishment of goals, the presence or absence of KR adding nothing, once differences due to past performance are accounted for. Unfortunately, KR and previous performance may have been confounded in Locke and Bryan's (1968) design. Higher goals were not stated until after the last trial. By that time, however, the KR *S*s had performed at a significantly higher level. Thus, it is impossible to tell whether the higher goals were a function of the KR manipulation or the higher performance since the timing of the goal estimates (after the last trial) makes this an ex post facto, not an experimental, study.

One might expect both variables to influence the level of goals set. Both, for example, presumably reflect on individual's abilities and motivational states (e.g., achievement motivation). Persons do not normally aim for goals that their previous performance indicates would be unrealistic or nonchallenging. Thus, previous performance would be expected to exert a direct influence on goal level. Knowledge concerning the favorableness or acceptability of past performance, however, is normally mediated by relevant others in one's environment. Thus, the effects of intentional manipulations of KR are also potentially important.

The present study was designed to investigate the impact of past performance and several KR conditions on the level of goals set. Two hypotheses were tested: (*a*) the higher the level of previous performance, the higher the level of goals set for future performance; and (*b*) when the effects of previous performance are accounted for, the greater the amount and accuracy of KR, the higher the level of goals set.

METHOD

Subjects

The *S*s were 80 undergraduate student volunteers (22 males, 58 females) enrolled in an introductory psychology course at the University of Wisconsin. The experiment was conducted in a regular classroom during four consecutive sessions. Each of the 80 *S*s had signed for one of the four designated sessions thus creating four groups of 20 each.

Task and Procedure

The study employed a simple addition task. Each *S* was given partially completed artificial copies of the monthly time sheets used by the hourly student assistants employed at the University. The daily totals had been entered in a column on the right-hand side of each sheet. The task involved totaling the column of daily hours to determine the monthly total and entering this total at the bottom of the column. The total number of hours worked during the month ranged from 139 to 159 with a mean of 148.

Each *S* was given 20 different time sheets repeated 4.5 times to make a total of 90. A pilot study indicated that *S*s would not be able to complete all 90 during the time allotted to the experiment. The time sheets were ordered so that each *S* in a group would have a unique combination that was matched by one equivalent combination in every other group.

To control competition, KR was provided only in numbers and not in relation to a group standard or average. Thus, each *S* had only his past performance to use as a standard for appraising KR in order to set goals. Talking was also prohibited during the experimental period and was monitored during the rest period while *E* was absent. These controls may not have been successful in entirely eliminating competition, however, since within each group *S*s could observe one another. The *S*s could have used these observations to establish a rough group standard for evaluating KR and for setting goals. However, there is no reason to expect differential effects of this type of possible contamination across the four conditions of the study.

In each group, *E* employed the following procedure to obtain Period-1 performance data. Each S was given a manila folder containing a blank time sheet, a blank orange-colored sheet, and the 90 time sheets, in that order. The *S*s were instructed to open the folder and look at the blank time sheet. The addition task was then explained. The *S*s were told to complete each time sheet in the order in which they appeared in the folder. At a given signal, *S*s removed the blank time sheet and orange-colored paper and started working on the time sheets. After 20 min. *S*s were asked to stop working on the task and to insert the orange-colored sheet after the last time sheet worked on whether or not it was completed.

The *E* then collected the folders and told *S*s that they would have a rest period of 5 min. During this period, *S*s were required to remain seated at their desks and talking was not permitted. An *O* monitored the groups as *E* took the folders into another room to tabulate the results. During the 5-min. period, *E* counted the number of completed time sheets in each folder and checked each time sheet against a correct listing of the total number of hours for each sheet. From this tabulating procedure a master sheet was prepared listing each *S* according to the number on his folder and giving the number of time sheets he completed and the number completed correctly. After the rest period, *E* returned to the experimental room and distributed the appropriate folders to each of the 20 *S*s.

At this point conditions in each of the four groups differed. Each *S* in the KR_C (correct) group was informed of the total number of time sheets he completed and the number completed correctly. The KR_I (incomplete) group was informed only of the number of time sheets completed incorrectly. (The *S* was not told the total number of time sheets that were completed.) The KR_N (no) group was given no information about individual performance. The KR_{EL} (erroeous low) group was informed that 50% of the number actually completed were correct.

After the KR conditions were differentially established, all groups were again given the same instructions. On a printed form *S*s were asked to indicate "How many time sheets do you expect to complete" in the second work period of 20 min. The latter data constituted the measure of goals employed in the present study. After working on the task for an additional 20 min., all time sheets were collected. The *S*s were then thanked for their participation in the study and dismissed.

RESULTS

The hypotheses stated earlier now can be elaborated as follows: (*a*) Goals are a positive function of previous performance; and (*b*) When the effects of previous performance are accounted for, highest goals will be set by the KR_C group, next highest goals will be set by the KR_I group, second lowest goals will be set by the KR_N group, and the lowest goals will be set by the KR_{EL} group.

The design employed one factor (KR) with four treatment levels (KR_C, KR_I, KR_N, KR_{EL}) and one continuous variable or covariate, Period-1 performance (P). Given the design employed and the hypotheses of interest, the data can be analyzed similarly using either analysis of variance/covariance (AV/ACV) or multiple regression (MR). The latter procedure was chosen because estimates of variance explained and appropriate partial correlations are a part of the normal MR output of the computer program available for our use. A reader who is more familiar with traditional AV/ACV is referred to Cohen (1968) for a discussion of the similarities between MR and AV/ACV procedures and output.

Analysis was first conducted to determine if KR alone was associated with variation in goals. For analytical purposes KR was categorized into three zero-one "dummy" variables with the KR_N control group coded zero in all three categories. Significance of the effect of KR on goals is provided by an *F* test of the multiple coefficient of determination (R^2) when goal level is the dependent variable and all three dummy variables are included as independent variables. R^2 indicates the proportion of goal variance attributable to KR and is exactly analogous in this situation to eta-square in analysis of variance. Moreover, given the procedure employed, the regression coefficients (b_i) associated with each KR_i condition indicate the mean goal-level difference between that KR group and the KR_N group (Cohen, 1968). As such, each b_i reflects the average impact of KR_i on goal levels relative to receiving KR_N.

Using the procedure described above, it was found that KR had a significant impact ($F = 5.18$; $p \leq .01$) on the level of goals set ($R^2 = .17$). Table 1 shows mean goal levels for all four KR groups and mean goal differences between each KR treatment and the control group (KR_N). The partial correlation coefficients show the relevance of these goal differences in correlational terms. It can be seen that only the KR_C group set significantly different (higher) goals than the KR_N group.

Since it was hypothesized that previous performance would influence the goal levels set,

TABLE 1

UNADJUSTED KR_C, KR_I, AND KR_{EL} VERSUS KR_N AS A DETERMINANT OF GOALS

Condition	Unadjusted *M*	Difference with KR_N (b_i)	Partial *r*	*t*
KR_N	14.85	—	—	—
KR_C	20.90	6.05	.36	3.38*
KR_I	15.15	.30	.02	.17
KR_{EL}	15.30	.45	.03	.25

Note.—Abbreviations: KR_N = no knowledge of results, KR_C = correct knowledge of results, KR_I = incomplete knowledge of results, KR_{EL} = erroneous low knowledge of results.
* $p < .01$.

additional analysis was conducted to identify this effect, if any, and to determine whether performance differences would mediate the KR effects observed in Table 1 where performance was not accounted for. In an MR context, performance is treated as a continuous independent variable introduced into the regression equation prior to KR; the latter continues to be categorized into three dummy variables. Using this procedure it is possible to identify: (*a*) the impact of performance on goal levels, (*b*) an overall incremental impact of KR on goals when performance is partialled out (exactly analogous to ACV treating performance as the covariate), and (*c*) mean differences between each KR and the KR_N groups after adjustment for any between-group performance differences.

Regression analysis showed that 26% of the goal variance was accounted for by the differences in previous performance ($F = 28.90$; $p < .01$). When KR was entered into the equation following performance, the amount of goal variance explained increased to .44 ($F = 19.80$; $p < .01$). The latter equation was employed to calculate adjusted mean goal levels (i.e., taking account of previous performance differences) and to test for differences between each KR treatment and KR_N. The results of this analysis are shown in Table 2. In addition, partial correlations are reported showing the effect of each KR condition relative to KR_N in correlational terms when performance is held constant. The results generally confirm the findings obtained when previous performance was not controlled. The KR_C group set significantly higher goal levels than the KR_N group. When previous performance is accounted for, the KR_{EL} group set lower goals than the KR_N group as hypothesized. However, neither KR_{EL} nor KR_I groups differed significantly from the KR_N group.

TABLE 2

ADJUSTED KR_C, KR_I, AND KR_{EL} VERSUS KR_N AS A DETERMINANT OF GOALS

Condition	Adjusted M	Difference with KR_N (b_i)	Partial r	t
KR_N	15.20	—	—	—
KR_C	20.74	5.54	.40	3.74*
KR_I	16.37	1.17	.09	.78
KR_{EL}	13.69	−1.51	−.11	−1.00

Note.—Abbreviations: KR_N = no knowledge of results, KR_C = correct knowledge of results, KR_I = incomplete knowledge of results, KR_{EL} = erroneous low knowledge of results.
* $p < .01$.

Thus, the results of the present study show that: (*a*) KR influences goal levels. (*b*) Correct KR increases goal levels significantly above those generated by KR_N, while KR_I and KR_{EL} do not. (*c*) Previous performance on the task affects goal levels, the greater the performance, the higher the subsequent goal. (*d*) When the effects of previous performance are accounted for, *KR* still influences goal levels. And (*e*) when the effects of previous performance are accounted for, correct KR still increases goal levels significantly above KR_N while KR_I results in insignificantly higher, and KR_{EL} results in insignificantly lower goal levels than KR_N.

DISCUSSION

It appears clear that KR influences the level of goals set for subsequent task performance. The relationship is, however, more complex than this conclusion implies. Previous performance also exerts a significant impact on level of goals. In the present study, performance and KR accounted for almost half of the goal variance. This finding is particularly important in view of the possible confounding effects of KR and performance on goals in the Locke and Bryan (1968) study. In that study, *S*s made goal statements three times; first, they were asked what their goals had been on the first eight trials after the eighth trial; second, they were asked what their goals were for the next eight trials; and third, they were asked (after the sixteenth trial) what their goals had been for the second eight trials. All *S*s were given knowledge of the correctness of their scores after each of the 16 trials. The KR *S*s kept track of their scores while the KR_N *S*s were not allowed to do so. Significant differences in performance between the KR and KR_N groups were found only for the second set of eight trials. It was also found that significant differences in goal statements were obtained only on the third goal statement questionnaire. Thus, at that point, the KR *S*s had performed at a higher level making it impossible to distinguish between KR and performance as determinants of goals. Since no differences in goals were found between the KR and KR$_N$ *S*s at the end of Trial 8 (where no significant performance differences had emerged), the suspicion might have been that past performance was more instrumental in determining goals than was KR. On the other hand, the present study more clearly demonstrates the independent and combined effects of KR and performance on goals.

It is interesting to note that approximately half of the variance was not explained by KR and performance. Additional possible determinants of goals might be the motivational states (need for achievement and competence) and aspiration levels of performers. As noted earlier, these characteristics may be reflected in past performance and goals, but only partially so. In addition, goal levels may be influenced by the performer's perceptions of the credibility of the task upon which he is asked to establish goals and the meaningfulness or value of the task to his own self-esteem and central interests. The influence of these potential determinants on goal levels were not investigated in this study.

In addition, all forms of KR do not influence goals to the same degree. In fact, when previous performance is accounted for, KR$_{EL}$ actually resulted in somewhat lower goals than KR$_N$. KR$_I$, even though accurate, increased goal levels only slightly above KR$_N$.

If replicated in field studies, the potential implications of these findings for administrative practice are somewhat disquieting. Managers are often urged to let employees know how they stand, that is, provide KR. If the standards against which performance is to be evaluated are not clearly defined and/or if the measurement processes used are not sufficiently specific to generate accurate KR, little may be gained by providing performance feedback. At best, incomplete feedback may have little impact on goals over providing no KR. A policy of providing conservative performance feedback (erroneous low) on the assumption that it is better to let the subordinate be surprised on the "upward side" may actually be less functional than providing no feedback.

REFERENCES

Adams, J. S. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. Vol. 2. New York: Academic Press, 1965.

Cohen, J. Multiple regression as a general data-analytic system. *Psychological Bulletin*, 1968, **70**, 426–443.

Cummings, L. L., & El Salmi, A. M. Empirical research on the bases and correlates of managerial motivation: A review of the literature. *Psychological Bulletin*, 1968, **70**, 127–144.

Herzberg F. *Work and the nature of man*. Cleveland: World Publishing, 1966.

Herzberg, F., Mausner, B., & Snyderman, B. B. *The motivation to work*. New York: Wiley, 1959.

Locke, E. A. Effects of knowledge of results, feedback in relation to standards, and goals on reaction-time performance. *The American Journal of Psychology*, 1968, **81**, 566–574. (a)

Locke, E. A. Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 1968, **3**, 157–189. (b)

Locke, E. A. What is job satisfaction? *Organizational Behavior and Human Performance*, 1969, **4**, 309–336.

Locke, E. A., & Bryan, J. F. Goal-setting as a determinant of the effect of knowledge of score on performance. *The American Journal of Psychology*, 1968, **81**, 398–406.

Locke, E. A., & Bryan, J. F. Knowledge of score and goal level as determinants of work rate. *Journal of Applied Psychology*, 1969, **53**, 59–65. (a)

Locke, E. A., & Bryan, J. F. The directing function of goals in task performance. *Organizational Behavior and Human Performance*, 1969, **4**, 35–42. (b)

Locke, E. A., Cartledge, N., & Knerr, C. S. Studies of the relationship between satisfaction, goal-setting, and performance. *Organizational Behavior and Human Performance*, 1970, **5**, 135–158.

Locke, E. A., Cartledge, N., & Koeppel, J. Motivational effects of knowledge of results: A goal-setting phenomenon? *Psychological Bulletin*, 1968, **70**, 474–485.

Porter, L. W., & Lawler, E. E. *Managerial attitudes and performance*. Homewood, Illinois: Dorsey, 1968.

Pritchard, R. D. Equity theory: A review and critique. *Organizational Behavior and Human Performance*, 1969, **4**, 197–211.

Schwab, D. P., & Cummings, L. L. Theories of performance and satisfaction: A review. *Industrial Relations*, 1970, **9**, 408–430.

Vroom, V. H. *Work and motivation*. New York: Wiley, 1964.

(Received September 2, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 531–536

EFFECTS OF CONTINGENT AND NONCONTINGENT REWARD ON THE RELATIONSHIP BETWEEN SATISFACTION AND TASK PERFORMANCE

DAVID J. CHERRINGTON

Graduate School of Business Administration, University of Illinois

H. JOSEPH REITZ[1] AND WILLIAM E. SCOTT, JR.

Graduate School of Business, Indiana University

It was proposed that there is no inherent relationship between satisfaction and productivity, and that relationships between the two variables are highly dependent upon performance-reward contingencies. Ninety *S*s performed a task for 1 hr. A monetary reward was then delivered to 21 of 42 high performers and 21 of 42 low performers. The *S*s next completed self-report measures of satisfaction and attitudes and performed the same task for another hour. Correlations between self-reports of satisfaction and second-hour productivity over all *S*s was .00. Significant positive correlations, however, were found between satisfaction and productivity of appropriately reinforced *S*s (rewarded high performers and nonrewarded low performers) while significant negative correlations were found for inappropriately reinforced *S*s (rewarded low performers and nonrewarded high performers).

Many current speculations on the relationship between worker satisfaction and task performance, as reviewed by Schwab and Cummings (1970), still imply that performance and satisfaction are causally related in one direction or the other. Some theorists have now added moderating variables to their behavioral formulas in hopes of facilitating the prediction of one variable from the other.

In contrast, the present authors postulate not only that (*a*) there is no inherent relationship between satisfaction and performance, but also that (*b*) one can produce about any empirical relationship between task performance and self-reports of satisfaction that one wishes. The first proposition is consistent with the conclusions of empirical reviews such as Brayfield and Crockett (1955) and with theoretical positions such as those of Porter and Lawler (1968). The second proposition was derived from operational proposals by Skinner (1969) and Bandura (1969) and from speculations by a variety of reinforcement theorists (e.g., Berlyne, 1967; Bindra, 1968; Rescorla & Solomon, 1967; and Weiskrantz, 1968).

Skinner (1969), for instance, insists that feelings are, at best, accompaniments rather than causes of behavior and that both are the products of common environmental variables. Bandura (1969, p. 598) pointedly suggests that we might better treat self-reports of satisfaction simply as another class of behavior rather than as indexes of an underlying state endowed with special causal powers. From this point of view, there exists no inherent relationship between performance and self-reports of satisfaction, and the empirical problem becomes that of examining the conditions under which the different response systems are correlated *or* independent.

We suggest that the type of reward system under which workers perform might strongly influence the satisfaction-performance relationship. One should be able to directly influence satisfaction by proferring or withdrawing rewards. To significantly influence performance, however, one must use *performance-contingent* rewards. Theoretically, then, by manipulating the contingencies of a reward system, one should be able to create conditions under which satisfaction and performance can be, empirically, either independent, or positively related, or negatively related.

To test these propositions let us define three types of performance-reward systems:

1. Random rewards. Rewards are distributed on bases independent of performance. The percentage of high performers receiving rewards does not differ from the percentage of low performers receiving rewards. Under this system we hypothesize that correlations between satisfaction and performance would not significantly differ from zero.

---

[1] Requests for reprints should be sent to H. Joseph Reitz, Business 650, Graduate School of Business, Indiana University, Bloomington, Indiana 47401.

2. *Positively contingent rewards*. Rewards are based directly on performance. All high performers are rewarded; all low performers are not rewarded. Under this system we hypothesize a positive correlation between satisfaction and performance.

3. *Negatively contingent rewards*. Rewards are based on factors inversely related to performance. Low performers are rewarded; high performers are not rewarded. Under this system we hypothesize a negative correlation between satisfaction and performance.

METHOD

Subjects

The *S*s were 90 undergraduate students, both male and female, enrolled in a junior-level business course. Students were enlisted as volunteers to score Closure Flexibility Tests for which they would be paid at least $1 per hour.

Procedure and Task

The *S*s reported to the laboratory in groups that varied from seven to nine individuals and were met by *E* who introduced himself as a graduate student who would be their supervisor.

When all *S*s had arrived and had familiarized themselves with the Closure Flexibility Test booklet, each was told that his task would be to score the tests at his work station. The *S*s were also told that the tests had been completed by employees in a paper mill. All of the Closure Flexibility Test booklets had in fact been marked according to 24 patterns of responses. Therefore, the difficulty of the task was controlled and the correct responses were known in advance.

The *S*s were told that they would be paid a minimum of $1 per hour but that the best performers in terms of quality and quantity would receive an additional $1 bonus. The *S*s were told that each had a 50–50 chance of receiving the bonus. The task of scoring the test booklets was explained to *S*s, aided by diagrams showing a partially marked answer sheet.

After *S*s began the task, *E* returned every 10 min. to bring additional booklets and collect the booklets and answer sheets that had been completed. At the end of 1 hr., *E* stopped *S*s and asked them to indicate on four 7-point scales an estimate of their quantity, quality, and overall performance and probability of receiving a reward.[2] The experimental manipulations of financial reward were then performed, which consisted of the following statements and payments.

> I have collected the tests as you have scored them and selected a sample from each of you to check your work. I've used a rather complicated index which combines quantity and quality of performance into one index. Based upon this index four of you will receive $2 for the last hour's work. The other four will get $1. The four winners are · · · and the four losers are · · ·.

In fact, the monetary bonus was randomly distributed to half of the high-performing and half of the low-performing *S*s.

The performance score for each *S* was the total number of rows of figures in the Closure Flexibility Test booklets that *S* scored correctly. This was determined by subtracting the number of errors from the total number of rows scored. Thus performance scores were a measure of both quality and quantity.

The *S*s were then asked to complete a self-report measure of satisfaction developed by Scott and his colleagues (Scott, 1967; Scott & Rowland, 1970). The format of the self-report measure was a semantic differential questionnaire as shown below.

Me At This Task

	Extremely	Quite	Slightly	Neither one nor the other	Slightly	Quite	Extremely	
Appreciated	___	___	___	___	___	___	___	Unappreciated
Bored	___	___	___	___	___	___	___	Interested
Efficient	___	___	___	___	___	___	___	Inefficient

Bipolar adjective pairs were set against four concepts: me at this task, my pay, my fellow workers, and the task. The response to each scale was scored from one to seven with seven assigned to that response which appeared to indicate the most preferred condition. A factor score was computed for each *S* by averaging *S*'s responses to each of the scales previously found to comprise that factor.

Eight factor scores were computed for each *S*. General affective tone score was obtained by averaging *S*'s responses to the following bipolar scales set against the concept, me at this task: appreciated-unappreciated, rewarded-penalized, satisfied-dissatisfied, and encouraged-discouraged. The remaining self-report measures and the semantic scales defining each factor were as follows—general arousal (me at this task): interested-bored, spirited-lifeless, and alert-listless; personal competence (me at this task): efficient-inefficient, productive-unproductive, reliable-unreliable, and effective-ineffective; general satisfaction with pay (my pay): pleasing-annoying, reasonable-unreasonable, superior-inferior, and rewarding-penalizing; equitableness of pay (my pay in comparison with what others in my group received): fair-unfair, high-low, and reasonable-unreasonable; adequacy of pay (my pay in comparison with what others get for similar work on the campus): superior-inferior, high-low, and reasonable-unreasonable; interpersonal attractiveness (my fellow workers): sociable-unsociable, helpful-obstructive, pleasant-unpleasant, unselfish-selfish, and cooperative-uncooperative; task attractiveness (the task): attractive-repulsive, exciting-dull, good-bad, interesting-boring, superior-inferior, and wholesome-unwholesome.

---

[2] There were no significant differences between the reward and nonreward group in either the subjective probability of being rewarded (5.42 vs. 5.22, respectively) or in estimated performance (5.37 vs. 5.32).

After completing the semantic differential scales, *S*s took a 5-min. break, after which they were given more test booklets and asked to continue scoring them for another hour. The procedure at the end of the second hour was identical to that at the end of the first hour. First, *S*s were asked to estimate their perceived performance and probability of reward on the four scales. Then they were paid and asked to fill out the semantic differential questionnaire. The monetary bonus was distributed to the same *S*s who received it at the end of the first hour.

After the questionnaires were filled out for the second time, *S*s were debriefed. They were told that all rewards were given at random, and since it was only by chance someone received more than the others, the winners were asked to share their rewards with the losers, although *E* did not insist that they do so.

RESULTS

We have hypothesized that there is no inherent relationship between satisfaction and performance, but that there is a direct relationship between the rewards and self-reports of satisfaction. Whatever covariation one does observe between satisfaction and performance, then, may depend upon other environmental conditions, particularly the kind of performance-reward contingency that has been arranged or has evolved.

The results summarized in Table 1 demonstrate the differential effects of rewards on performance and satisfaction.[3] Performance scores of rewarded *S*s did not differ from those of nonrewarded *S*s.

Mean satisfaction scores of rewarded *S*s, however, were significantly greater than mean scores of nonrewarded *S*s on each of the eight satisfaction indexes. Thus self-report measures of satisfaction can be highly dependent on the occurrence or nonoccurrence of monetary reward, but if that reward is delivered independent of performance, it fails to have incremental effects on subsequent performance.

To test the effects of variations in performance-reinforcement contingency, *S*s were classi-

TABLE 1

COMPARISON OF MEAN PERFORMANCE AND SATISFACTION SCORES OF REWARDED VERSUS NONREWARDED SUBJECTS

Item	$\bar{X}$ Scores		D_m
	Rewarded *S*s	Nonrewarded *S*s	
Performance			
First hour	285.6	286.6	−1.0
Second hour	403.8	411.2	−7.4
Satisfaction indexes			
General affective tone	5.16	3.83	1.33***
General arousal	4.98	4.44	.54**
Personal competence	5.82	5.01	.81***
General satisfaction with pay	5.61	3.69	1.92***
Equity of pay	5.24	3.74	1.50***
Adequacy of pay	4.76	3.40	1.36***
Attractiveness of fellow workers	4.96	4.63	.33*
Attractiveness of task	3.80	3.19	.61***

Note.—Abbreviation D_m = Difference of means.
* $p < .01$.
** $p < .005$.
*** $p < .001$.

---

[3] Raw scores are reported in the tables. Normalized scores, however, were used in all statistical analyses.

TABLE 2

COMPARISON OF MEAN PERFORMANCE AND SATISFACTION SCORES OF APPROPRIATELY VERSUS INAPPROPRIATELY REINFORCED SUBJECTS

Item	$\bar{X}$ Scores		D_m
	Appropriately reinforced *S*s	Inappropriately reinforced *S*s	
Performance			
First hour	291.6	280.9	10.7
Second hour	428.4	385.9	42.5**
Satisfaction			
General affective tone	4.26	4.11	.15
General arousal	4.12	4.02	.10
Personal competence	5.30	5.07	.23
General satisfaction with pay	4.56	4.11	.45
Equity of pay	4.40	4.18	.22
Adequacy of pay	4.18	3.64	.54*
Attractiveness of fellow workers	4.85	4.78	.07
Attractiveness of task	3.31	3.27	.04

Note.—Abbreviation D_m = Difference of means.
* $p < .05$.
** $p < .01$.

fied on the basis of first-hour performance. The 42 *S*s whose performance scores were above the median of their group are referred to as high performers and the 42 *S*s below the median as low performers. Six *S*s who performed at the median were excluded from further analysis. Rewards at the end of the first hour were distributed randomly among high and among low performers so that half the *S*s in each performance classification were rewarded. The appropriateness of each *S*'s reinforcement was then defined by the performance-reward contingency that happened to occur for that *S*. The 21 rewarded high performers and the 21 nonrewarded low performers were classified as appropriately reinforced. The 21 rewarded low performers and the 21 nonrewarded high performers were classified as inappropriately reinforced.

TABLE 3

CORRELATIONS BETWEEN SATISFACTION AT END OF FIRST HOUR AND SUBSEQUENT PERFORMANCE DURING SECOND HOUR

Satisfaction indexes	All *S*s[a]	Appropriately reinforced *S*s[b]	Inappropriately reinforced *S*s[b]
General affective tone	.00	.56***	−.32*
General arousal	.10	.42**	−.15
Personal competence	.21	.54***	−.01
General satisfaction with pay	.04	.46**	−.29
Equity of pay	−.17	.15	−.44**
Adequacy of pay	.04	.39**	−.31*
Attractiveness of fellow workers	.22*	.33*	.13
Attractiveness of task	.04	.21	−.08

[a] $n = 90$.
[b] $n = 42$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

The results shown in Table 2 reveal, first, that the average second-hour (postreinforcement) performance of appropriately reinforced *S*s was significantly higher than that of inappropriately reinforced *S*s, whereas the first hour (prereinforcement) performance of the two classes did not differ significantly. Comparisons between the two classes on self-reports of satisfaction, however, revealed a significant difference on only one of the eight satisfaction indexes, adequacy of pay. The tests summarized in Tables 1 and 2 thus establish that, as hypothesized, it is the contingency of a reward on performance that enables it to bring about increments in performance. That contingency characteristic, however, has little or no effect on satisfaction.

Two different relationships between satisfaction and performance were investigated: (*a*) the relationship between self-reports of satisfaction and subsequent performance, and (*b*) the relationship between performance and subsequent self-reports of satisfaction.

Table 3 presents the results of correlational analyses performed between the eight satisfaction indexes recorded at the end of the first hour and subsequent performance during the second hour. The data for all *S*s, among whom as many low performers as high performers were rewarded, revealed only one significant correlation between satisfaction and subsequent performance. However, when the data were separated according to *S*s' performance-reward contingencies, distinct differences in the satisfaction-performance relationships were observed. For the 42 *appropriately* reinforced *S*s, significant *positive* correlations were found between six of the eight satisfaction indexes and subsequent performance. But for the 42 *inappropriately* reinforced *S*s, *negative* correlations were found between seven of the eight satisfaction indexes and subsequent performance, of which three correlations were statistically significant, and one closely approached significance at the .05 level.

Table 4 presents the results of correlational analyses performed between performance during the second hour and the eight satisfaction indexes subsequently recorded at the end of the second hour. The similarities between Table 3 and 4 are striking. The data for all *S*s revealed no significant correlations between performance and subsequent satisfaction. However, when distinctions were made between appropriately and inappropriately reinforced *S*s, where the performance classifications were based on the second hour's productivity scores, widely disparate performance-satisfaction relationships were observed. For the appropriately reinforced *S*s, significant positive correlations were obtained between performance and each of the eight satisfaction indexes. For the inappropriately reinforced *S*s, negative correlations were obtained between performance and seven of the eight satisfaction indexes, with four of the correlations being highly significant.

TABLE 4

CORRELATIONS BETWEEN PERFORMANCE DURING SECOND HOUR AND SATISFACTION AT END OF SECOND HOUR

Satisfaction indexes	All *S*s[a]	Appropriately reinforced *S*s[b]	Inappropriately reinforced *S*s[b]
General affective tone	−.03	.55***	−.51***
General arousal	.02	.42**	−.26
Personal competence	.13	.48**	−.16
General satisfaction with pay	.03	.67***	−.56***
Equity of pay	−.09	.45**	−.51***
Adequacy of pay	−.03	.59***	−.57***
Attractiveness of fellow workers	.20	.44**	.04
Attractiveness of task	−.06	.32*	−.16

[a] $n = 90$.
[b] $n = 42$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

DISCUSSION

The results of this study support the hypothesis that the nature and magnitude of the relationship between satisfaction and performance depend heavily upon the performance-reward contingencies that have been arranged. The findings are consistent with the predictions of the Porter and Lawler (1968) model, although the theoretical bases for our hypothesis differ from that model. For instance, Porter and Lawler's model implies that, under contingent-reward conditions, performance causes satisfaction because performance leads to rewards which, in turn, cause satisfaction. Our theory implies no cause-effect relationship between performance and satisfaction; instead, it stresses the performance-reinforcing as well as the satisfaction-increasing potential of contingent reinforcers.

The random reward system described here has as its organizational counterpart companies in which rewards such as pay increases and promotion are alleged to be performance-contingent but, in fact, are contingent upon factors independent of performance, such as age, seniority, or upward influence of one's supervisor. In such an organization we would expect satisfaction and performance to be unrelated.

The negatively contingent reward system is analogous to an organization in which rewards are based on factors that happen to be inversely related to performance. Such systems are not, unfortunately, as rare as one might think. One of the authors recently encountered a company in which pay increases were based on supervisor ratings, until it was discovered that the supervisors' ratings were negatively related to performance. In such an organization we would expect to find satisfaction and productivity negatively related. High performers are dissatisfied because they feel cheated; low performers are more than satisfied with their relative increases in pay.

So the importance of taking into account the contingencies between performance and rewards is again emphasized for those who seek to affect employee attitudes and/or behavior through various reward systems. If rewards are not positively contingent, then the administration of rewards will not only fail to encourage performance increments, it may also increase dissatisfaction and, ultimately, absenteeism and resignations among the highest producing employees.

REFERENCES

BANDURA, A. *Principles of behavior modification.* New York: Holt, Rinehart & Winston, 1969.

BERLYNE, D. E. Arousal and reinforcement. In D. Levine (Ed.), *Nebraska symposium on motivation: 1967.* Lincoln: University of Nebraska Press, 1967.

BINDRA, D. Neuropsychological interpretation of the effects of drive and incentive-motivation on general activity and instrumental behavior. *Psychological Review*, 1968, **75**, 1–22.

BRAYFIELD, A. H., & CROCKETT, W. H. Employee attitudes and employee performance. *Psychological Bulletin*, 1955, **52**, 396–424.

PORTER, L. W., & LAWLER, E. E., III. *Managerial attitudes and performance.* Homewood, Illinois: Irwin-Dorsey, 1968.

RESCORLA, R. A., & SOLOMON, R. L. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychological Review*, 1967, **74**, 151–182.

SCHWAB, D. P., & CUMMINGS, L. L. Theories of performance and satisfaction: A review. *Industrial Relations*, 1970, **9**, 408–430.

SCOTT, W. E., JR. The development of semantic differential scales as measures of "morale." *Personnel Psychology*, 1967, **20**, 179–198.

SCOTT, W. E., JR., & ROWLAND, K. M. The generality and significance of semantic differential scales as measures of "morale." *Organizational Behavior and Human Performance*, 1970, **5**, 576–591.

SKINNER, B. F. *Contingencies of reinforcement: A theoretical analysis.* New York: Appleton-Century-Crofts, 1969.

WEISKRANTZ, L. Emotion. In L. Weiskrantz (Ed.), *Analysis of behavioral change.* New York: Harper & Row, 1968.

(Received August 17, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 537–542

BACKGROUND CHARACTERISTICS AND WORK VALUES DISTINGUISHING SATISFACTION LEVELS AMONG ENGINEERS[1]

DONALD A. WOOD[2]

Graduate School of Business, Indiana University

This study attempted to determine if those with higher job satisfaction in a national sample of 1,500 professional engineers could be distinguished from those with lower levels of satisfaction on the basis of personal, background, and structural variables. Linear discriminant function analysis disclosed those most satisfied were best discriminated with 67% and 65% classification accuracy in initial and replication groups, respectively. Discrimination for those least satisfied was second best with 63%—47% hit rates. Group discriminators were best described by professional training and knowledge, job values, and participation in technical activities. An adapted method for applying function weights to individual cases was described.

Individuals within different occupations, skill classifications, and organizational levels often indicate marked differences in job satisfaction (e.g., Centers & Bugental, 1966; Doll & Gunderson, 1969; Porter, 1962). In addition, demographic variables like age or education have been significantly correlated with satisfaction measures, implying subgroups determined by such variables also vary on job satisfaction (Fournet, Distefano, & Pryer, 1966). Little research, however, has addressed itself to determining the characteristics that best distinguish workers grouped according to how satisfied they are with their work.

A fundamental question emerges as to what background or structural variables can be identified that contribute most to the discriminability between groups with high and low job satisfaction, using the job satisfaction measure itself as a classifier. The more traditional interest regarding the identification of attitudinal correlates is replaced by the desire to seek out attitudinal discriminators instead.

Ronan (1970) has extensively reviewed research using a continuous satisfaction measure for correlational analysis *within* employee groups. These efforts mainly involve simple and multiple criteria-satisfaction correlations. Such results are important, but since regressions across different employee groups vary considerably, conclusions concerning attitudinal correlates within one group cannot be generalized to other groups. If the question concerns which variables can best distinguish *between* groups differing by the degree of overall job satisfaction, then discriminant function analysis, instead of regression analysis, becomes appropriate. The more customary concern of satisfaction regression within groups is replaced by the concern for selecting variables that best preserve the identity of groups differing in satisfaction. The discriminant function approach would appear capable of providing a clearer picture of the optimum combination of variables that differentiate such groups.

The purpose of this study is to determine what background, structural, and demographic items most distinguish various satisfaction levels among engineers and to identify an optimal weighting scheme by which such levels can be maximally discriminated.

METHOD

Sample

The sample consisted of 1,500 graduate engineers randomly selected from a larger national survey of engineers in industry and government (LeBold, Perrucci, & Howland, 1966). Seventy-one percent had bachelor's degrees, 17% had master's degrees, and 12% had doctorate degrees—all in various engineering fields. Eighty-five percent of the sample were employed in industry, with chemical, electrical, and aircraft companies highly represented. The remaining included engineers in Federal government (5%); state and local government (8%); and nonmanufacturing, including research and testing laboratories (2%). Three-fourths of the sample worked in organizations exceeding 1,000 employees, but their work unit size remained small, usually less than 10. The total sample had a median age of 39.

---

[1] This study was prepared under a grant from the Office of Manpower Policy, Evaluation and Research, Grant No. 91-16-68-36, United States Department of Labor under the authority of Title I of the Manpower Development and Training Act of 1962. The author is indebted to W. K. LeBold for permitting access to the data used in this study.

[2] Requests for reprints should be sent to Donald A. Wood, Graduate School of Business, Indiana University, Bloomington, Indiana 47401.

Procedure

Preceding the actual analyses, groups of engineers differing in job satisfaction had to be identified. This was done by summing across 60 job statements concerning areas such as autonomy, recognition, challenge, security, status, and interpersonal relations, all of which had been responded to on a 3-point scale—very characteristic (3), some characteristic (2), and not characteristic (1). For example, an engineer would have to indicate how characteristic is "using your abilities in challenging work" with respect to his current job. He responded in like manner to the remaining items. After adding across all 60 job items, a general attainment index resulted (range of 60–180). This is a variation of the Porter (1962) index obtained by asking "how much is there now?" and reflects the satisfaction summation process described by Evans (1969, p. 95). The rationale for such a score is that as work attainment or perceived characteristics increase, job satisfaction will also increase. This relationship has been previously demonstrated and supported (Wood & LeBold, 1970).

After the characteristic scores were calculated, the score distribution was divided into fifths, resulting in five different groups ranging from those who gave their job a low characteristic score (dissatisfied) to those who saw their position worthy of a high score (satisfied). Score intervals were determined from the total sample using the BMDO2S contingency table analysis program (Dixon, 1968, pp. 341–357) under the assumption that the distribution was normal in form. The resulting *N*s within the interval groupings used indicated the distribution was very nearly normal. After the total sample record tape was ordered by the characteristic score, it was decided to construct two separate data tapes, the second for replication purposes, each with 750 records and each with 150 *S*s in each of five groups. If an *S* from the total sample record tape was classified in Group 1, Tape 1, then the next *S* was assigned Group 1, Tape 2, etc. Both tapes were ordered according to the characteristic score magnitude so that group membership would be sequential.

The groups for Sample 1 are designated as follows: low (L1) for those whose characteristic scores are 116 or below; low middle (LM1) 117–127; middle (M1) for scores from 127–136; high middle (HM1) for scores ranging from 136–145; high (H1) for those scoring 145 or above. Comparable Sample 2 groups using the same intervals are designated L2, LM2, M2, HM2, and H2, respectively.

Program BMDO5M, linear discriminant analysis for several groups (Dixon, 1968, pp. 196–207), was used for the analyses.[3] Twenty items were selected for the discriminant analysis, primarily on the basis of their correlation with the characteristic score. However, if this selection process did not include items with known relationships to job satisfaction, as determined by previous research, they were included to the exclusion of lower correlated items.

The variable set selected as potential discriminators of job satisfaction included: (*a*) the departmental size currently employed in; (*b*) number of employees responsible for, both professional and nonprofessional; (*c*) hours per week spent at home on work-related activities; (*d*) amount of work done that could be done by someone else with less technical training; (*e*) extent of employer's continuing education opportunities; (*f*) extent of formal training in a variety of specific technical subjects and detailed engineering activities, and the engineer's state of knowledge in each area ranging from extensive to none; (*g*) organizational size; (*h*) degree of supervisory responsibility; (*i*) degree of technical responsibility; (*j*) a description of job function ranging from entirely technical to entirely administrative; (*k*) relevancy of engineering background to present job from "must have" to "unnecessary"; (*l*) a direct, overall, single item job satisfaction scale ranging from very satisfied to very dissatisfied; (*m*) number of memberships in national professional and scientific societies; (*n*) number of technical articles written for publication; (*o*) number of technical activities like meetings and outside courses engaged in during past year; (*p*) gross annual income; (*q*) peak gross annual income expected during career; (*r*) seriousness of doubts about engineering after having made career choice; (*s*) age; and (*t*) a summated importance score (very important, 3; some importance, 2; no importance, 1) estimated across the same 60 job statements used for calculating the attainment index. This latter score was an attempt to measure the extent to which various aspects of one's present job were personally valued by the engineer.

All 750 engineers comprising Sample 1 were used in the first discriminant analysis employing all 20 items as discriminators. Following the 20 variable analysis, the eight best discriminating items in Sample 1 were used again in a second analysis in Sample 2. These identifications were made by multiplying the normalized discriminate vector entries for each group by the square root of the corresponding diagonal element in the pooled within-group dispersion matrix (Cooley & Lohnes, 1962, p. 118). The resulting weights give the relative discriminating power for each variable in each discriminant function.

RESULTS

The scaled discriminant functions computed for each group in the 20-item, Sample 1 analysis are given in the first part of Table 1. The

---

[3] It may be well to distinguish between the linear discriminant function to be employed in this study and the linear regression function more commonly used in predictive studies. A linear regression function makes use of continuous dependent variable values for determining function weights to estimate other dependent values. The discriminant function weights come about due to classifications made on the dependent measure, not to continuous values. The "prediction" is group membership, not criterion value as in regression.

TABLE 1

FIVE SCALED DISCRIMINANT FUNCTIONS FOR FIVE GROUPS USING TWENTY ITEMS (SAMPLE 1) AND EIGHT ITEMS (SAMPLE 2)

Item	Scaled discriminant functions									
	Sample 1					Sample 2				
	L1	LM1	M1	HM1	H1	L2	LM2	M2	HM2	H2
1. Department size	.12	.10	.11	.11	.11					
2. No. employees responsible for	.14	.16	.17	.15	.19					
3. Hours of homework	−.06	−.06	−.06	−.05	−.05					
4. *Do work others should do*	.40	.41	.41	.39	.38	.36	.34	.32	.31	.30
5. Employer's policy on education	.27	.30	.32	.33	.32					
6. *Training and state of knowledge*	1.13	1.13	1.14	1.14	1.14	1.37	1.43	1.38	1.40	1.38
7. Size of organization	.09	.10	.10	.10	.10					
8. Supervisory responsibility	.18	.18	.19	.19	.18					
9. Technical responsibility	.18	.20	.20	.20	.20					
10. *Technical-administrative function*	.34	.34	.35	.39	.37	.58	.60	.61	.62	.63
11. *Relevance of background*	.40	.41	.40	.40	.39	.57	.54	.54	.53	.52
12. Overall job satisfaction	.25	.20	.16	.15	.12					
13. No. professional societies	−.30	−.31	−.31	−.31	−.34					
14. Articles written	.06	.07	.06	.07	.10					
15. *Technical activities*	.49	.51	.50	.48	.51	.91	.95	.93	.94	.96
16. Annual income	.25	.25	.22	.25	.25					
17. Expected income	−.10	−.11	−.08	−.11	−.11					
18. *Doubts about engineering*	.40	.40	.41	.42	.42	.29	.31	.33	.33	.33
19. *Age*	.44	.45	.46	.45	.44	.43	.44	.44	.44	.45
20. *Importance score*	.75	.80	.83	.87	.92	1.08	1.12	1.15	1.17	1.24

Note.—Items italicized have eight highest scaled function values across all five functions in Sample 1. Group *N*s = 150 Abbreviations: L = Low, LM = low middle, M = middle, HM = high middle, H = high.

weightings resulting from scaling are relative indexes of discriminatory efficiency and should not be confused with the original weights estimated in the discriminant functions. In this and successive analyses, Functions 1–5 are related to Groups L1(2) through H1(2), respectively. The item concerning training and state of knowledge (Item 6) appears to be the most effective single discriminator between the five satisfaction groups of engineers after all relative weightings in the scaled functions are compared. The importance score received the second highest weighting and the item pertaining to technical activities was third. It is noteworthy that weightings, in general, do not change drastically from one function to another.

Other notable discriminants included age, doubts about engineering as a career, doing work that could be done by others, technical-administrative function, and relevance of engineering background.

The classification matrix resulting from the first analysis is given in Table 2. This matrix actually summarizes the results of estimating five probability densities, one from each discriminant function, for all *S*s in the sample after which the engineers are assigned to the group in which the probability density is greatest. For example, the standardized raw score vector for S_1, Group 1, is multiplied, in turn, by each discriminant function. The result is five discriminant scores from which probability densities are determined (for computational procedure, see Dixon, 1968, p. 202). If the highest density occurs with Function 1, then S_1 is assigned to Group 1. If Function 3 gives the greatest density, this *S* is then assigned to Group 3, etc. More simply, each individual's response vector is compared to the mean response vector for each group. A person is assigned to the group where this difference is smallest. If all five group functions discriminated perfectly, the

TABLE 2

CLASSIFICATION MATRIX FOR FIVE GROUPS IN SAMPLE 1 USING TWENTY DISCRIMINANT ITEMS

Group	Discriminant function						% Group hits
	1	2	3	4	5	Total	
L1	*95*	21	15	12	7	150	.63
LM1	47	*37*	33	15	18	150	.25
M1	22	26	*53*	27	22	150	.35
HM1	17	11	32	*55*	35	150	.37
H1	6	10	12	21	*101*	150	.67
Total	187	105	145	130	183	750	
% Function hits	.51	.35	.37	.42	.55		

Note.—Chi-square value for table of 393.68 with $df = 16$ is significant at $p = .001$. Group Ns = 150. Italicized values represent the number of hits obtained from each discriminant function. Abbreviations: L = low, LM = low middle, M = middle, HM = high middle, H = high.

matrix diagonals would all be 150, and all other entries would be zero.

The italicized diagonal entries in Table 2 indicate the number of hits or correct classifications obtained from each discriminant function. The group hit percentage is that proportion of the total group accurately placed. The greatest hit rate is 67% for the most satisfied group, H1. This is followed by a 63% rate for the least satisfied group, L1. Group LM1 members were discriminated least effectively with only one-fourth placed in this group by its respective function. Just over one-third of those in Groups M1 and HM1 were considered hits. From these findings, it would appear that very satisfied and very dissatisfied engineers can be categorized rather accurately without knowing their actual attitude levels.

Another way of analyzing discriminant efficiency is to calculate the percentage of highest densities determined by a function that actually falls in the grouping associated with that function. This was done by dividing the column totals (see Table 2) into the corresponding diagonal entry giving a percentage of function hits. The highest functional percentage of 55 indicates that of 750 engineers in Sample 1, 183 received their greatest density estimate from Function 5 of which 101 or 55% were actually members of Group H1. It can be seen from Table 2 that the pattern of function hits follows that for groups (i.e., Function 5 associated with Group H1 is most effective in discriminating, and Function 2 for Group LM1 is least effective). It seems reasonable that the number of group hits would be reflected by the discriminant functions' efficiency index.

In order to apply a rather stringent replication of the discriminant analysis done in Sample 1, the eight highest-weighted (determined by the scaled functions) items from Sample 1 were used in a Sample 2, 8-item discriminant problem. Table 1 gives the scaled functions resulting from this analysis. With the exception of Item 1, doing work of others, and Item 6, doubts about engineering, all the relative discriminant weights are the same or higher here than in the Sample 1 analysis. It is most noteworthy, however, that training and state of knowledge, job values, and technical activities again received the highest weightings, in that order.

The hits percentage for Group H2 of 65 (see Table 3) was only slightly lower than the H1 rate of 67% when all 20 items were used. The next highest rate of 47% for Group L2 reflects a decline in discriminability that also occurred in Groups M2 and HM2. Considerably better discrimination occurred in Group LM2 using eight items than in LM1 using 20 items.

Despite the much higher hit rate for Group H2 when compared with Group L2, the corresponding function hit percentages are surprisingly similar. This is due to the large number of non-H2 designations made by Function 5 and the comparatively fewer non-L2 assignments made by Function 1. This finding would seem to indicate that the

TABLE 3

CLASSIFICATION MATRIX FOR FIVE GROUPS IN SAMPLE 2 USING EIGHT MOST-WEIGHTED DISCRIMINANT ITEMS FROM SAMPLE 1

Group	Discriminant function						% Group hits
	1	2	3	4	5	Total	
L2	*71*	34	13	12	20	150	.47
LM2	44	*48*	16	17	25	150	.32
M2	27	32	*22*	31	38	150	.15
HM2	21	28	23	*29*	49	150	.19
H2	9	15	13	16	*97*	150	.65
Total	172	157	87	105	229	750	
% Function hits	.41	.31	.25	.28	.42		

Note.—Chi-square value for table of 185.78 with $df = 16$ is significant at $p = .001$. Group Ns = 150. Italicized values represent the number of hits obtained from each discriminant function. Abbreviations: L = low, LM = low middle, M = middle, HM = high middle, H = high.

hit rate for engineering groups widely divergent in overall job satisfaction does not necessarily reflect the classifying efficiency of the corresponding group functions. From this discriminant analysis using the reduced data set, classification in the high and low-moderate groups does not suffer appreciably, while the hit rate for those least satisfied is reduced but still remains substantial. Data reduction is most severe on those in the middle and high-middle groups with their hits percentage going below that expected by chance (.20).

DISCUSSION AND CONCLUSIONS

It may be concluded from this study that five engineering groups differing in general job satisfaction as measured by a perceived work attainment index can be significantly distinguished by work value and background variables, particularly if satisfaction or dissatisfaction is extreme. Linear discriminant function analysis has demonstrated that item weights can be derived such that extreme satisfaction group identities can be reasonably maintained or approximated. Lowered classification success found in the three intermediate satisfaction groups could well have been due to the abbreviated satisfaction ranges used in their formation. Had they been merged into one large moderate group, discrimination might well have increased. But this would have been accomplished by sacrificing the distinctiveness of the classification scheme involved. Perhaps, more practically, the extremes of job satisfaction are most crucial if discrimination is desired. Behaviorally, it may make a greater difference if one is extremely happy or unhappy with his work as opposed to being mildly satisfied.

Since it has been reliably established that engineering attitude-group differences do exist with respect to professional training, values, and activities, the question of using the resulting discriminant weights arises. Stinson (1958) suggests a procedure that is adapted for use here. For illustration purposes, the three best discriminators are used in determining function scores (V) for the L, M, and H Sample 2 groups. Scaled item weights from the 8-item analysis functions and respective group means (training, values, and activities, in that order) are used in the following equations:[4]

$$V_{L2} = 1.367(189.250) + 1.082(134.510) + .911(16.830) = 419.57$$

$$V_{M2} = 1.382(190.220) + 1.147(141.490) + .934(17.530) = 441.54$$

$$V_{H2} = 1.382(192.740) + 1.239(152.310) + .959(18.250) = 472.57$$

The midpoints between these critical V scores can be used in the assignment of individual cases. For example, if a person (S_1) had raw scores on the training, importance, and activities items of 175, 138, and 17, respectively, his V score of 418.36 would place him far below the L2-M2 midpoint score of 430.55 and very close to the V score for L2 ($V_{S_1} = 1.391(175) + 1.152(138) + .938[17]$). Thus, he would be assigned to the least satisfied group, L2. The discriminant weights used with raw scores are the averaged scaled values across the five functions. The slightly increased precision gained from applying every function in turn to each raw score vector and then calculating probability densities for each does not seem justified. This adaptation of Stinson's (1958) method appears to be more practical and feasible with respect to assigning new *S*s to satisfaction groups.

---

[4] For illustration purposes, only the three best discriminators from the 8-item analysis are used. More definitive classification for questionable cases could be achieved by successively adding discriminators to the equation, much like adding predictors to a regression equation. A complete table of item means and standard deviations can be obtained from the author. (See address in Footnote 2.)

It should be emphasized that data other than the V score should be taken into account, especially in borderline cases. But the discriminant V score does help to integrate the predictions that come from the three most valid variables in these samples into a single number.

This classification scheme offers a workable method for placing *S*s in job satisfaction categories when their job satisfaction is unknown. More research is needed, however, to determine if state of training, work values, previous participation in professional activities, and other pertinent data gathered at the time of hire could designate initial groupings that would later be validated by follow-up attitude surveys. If so, discriminant analysis could serve to identify those with propensities to be satisfied or dissatisfied with their job prior to hire and placement.

REFERENCES

CENTERS, R., & BUGENTAL, D. E. Intrinsic and extrinsic job motivations among different segments of the working population. *Journal of Applied Psychology*, 1966, **50**, 193–197.

COOLEY, W. W., & LOHNES, P. R. *Multivariate procedures for behavioral sciences*. New York: Wiley, 1962.

DIXON, W. J. (Ed.) BMD biomedical computer programs. Berkeley, University of California Press, 1968. (No. 2)

DOLL, R. E., & GUNDERSON, E. K. E. Occupational group as a moderator of the job satisfaction-job performance relationship. *Journal of Applied Psychology*, 1969, **53**, 359–361.

EVANS, M. G. Conceptual and operational problems in the measurement of various aspects of job satisfaction. *Journal of Applied Psychology*, 1969, **53**, 93–101.

FOURNET, G. P., DISTEFANO, M. K. JR., & PRYER, M. W. Job satisfaction: Issues and problems. *Personnel Psychology*, 1966, **19**, 165–183.

LEBOLD, W. K., PERRUCCI, R., & HOWLAND, W. E. The engineer in industry and government. *Journal of Engineering Education*, 1966, **56**, 237–259.

PORTER, L. W. Job attitudes in management: I. Perceived deficiencies in need fulfillment as a function of job level. *Journal of Applied Psychology*, 1962, **46**, 375–384.

RONAN, W. W. Individual and situational variables relating to job satisfaction. *Journal of Applied Psychology*, 1970, **54** (1, Pt. 2).

STINSON, P. J. A method for counseling engineering students. *Personnel Guidance Journal*, 1958, **37**, 294–295.

WOOD, D. A., & LEBOLD, W. K. Overall job satisfaction and engineering job values, characteristics, and behavior. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 1970, **5**, 585–586. (Summary)

(Received July 29, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 543–548

ANXIETY AS A FUNCTION OF INTRINSIC-EXTRINSIC JOB ORIENTATION, THE PRESENCE OR ABSENCE OF OBSERVERS, AND TASK DIFFICULTY[1]

SHOUKRY D. SALEH[2]

Department of Management Sciences, University of Waterloo

Ninety-six intrinsically job-oriented and 96 extrinsically job-oriented male university students were divided into four comparable subgroups. Two subgroups worked on simple anagrams, one alone and one observed by two individuals of higher status. The other two worked on difficult anagrams under the same two conditions. Felt anxiety and perception of task difficulty ratings were elicited along with reactions toward the experimental conditions. The intrinsically oriented expressed more uneasiness than the extrinsically oriented while being observed on the difficult anagrams. They also tended to perceive the task as more difficult when observed than when alone, while the opposite was true for the extrinsically oriented. The importance of job orientation in the reaction to the presence or absence of observers was discussed.

The objective of this study is to clarify experimentally how intrinsically oriented and extrinsically oriented individuals would react to the presence of higher status observers. Intrinsically oriented individuals are those who would emphasize the actual performance of the job. They, for instance, would be concerned with the challenge that their job offers and with their accomplishment on it. Extrinsically oriented individuals, however, are those who would emphasize the environment in which the job would be performed. They would be more concerned with such factors as security, working conditions, and supervision than with challenge and accomplishment.

The reviews by Cummings and El-Salmi (1968), House and Wigdor (1967), and Whitsett and Winslow (1967) show that the literature has concentrated on supporting or refuting the proposition that the intrinsic and the extrinsic categories are two dimensions of job attitude. Very few studies, however, have dealt with their dynamics. Weissenberg and Gruenfeld (1968) for example, found a positive relationship between satisfaction with the intrinsic factors and job involvement. In another correlational study, Saleh and Grygier (1969) found that the intrinsically oriented *S*s had a stronger tendency to be independent and felt that they needed less guidance or reassurance than the extrinsically oriented ones. It has been also pointed out (e.g., Herzberg, 1966; Saleh, 1964) that underlying the intrinsic orientation is the need for self-actualization, while physiological, security, and acceptance needs are the basis of the extrinsic orientation.

The main interest in the audience effect literature has been concern with the effects of the presence of others on performance. Zajonc (1965), in an attempt to explain the opposing findings in this area, proposed that the presence of others has drive-arousing effects on those who are performing, and this arousal increases the probability of occurrence of the responses that are dominant in a person's repertory. Thus, if the task was a simple one in which the person has almost learned the solution, the correct response would be dominant and the presence of others would facilitate performance. On the other hand, if the task was complex and the correct response was not dominant in the hierarchy of possible responses, the presence of others would impede performance.

In his proposition, Zajonc indicated the mediating arousal effects only for the presence of others. The situation as a whole, however, should be considered a source of arousal, and in this case the nature of the task and its difficulty should not be ignored. That task difficulty is related to arousal has been demonstrated by Hess and Polt (1964) and by

---

[1] This research was supported by Grant No. 69-0141 from The Canada Council. The author wishes to thank M. Brown for his comments on an earlier draft and D. McNeil for his help in this research.

[2] Requests for reprints should be sent to Shoukry D. Saleh, Department of Management Sciences, University of Waterloo, Waterloo, Ontario, Canada.

Schnore (1959). Therefore, assuming that there is a curvilinear relationship between the level of arousal and performance (Hebb, 1955), it is possible that the arousal induced by a difficult or complex task added to that of audience presence may result in a general arousal level that exceeds the level for optimum or maximum performance. In the case of the simple task, however, the arousal the task contributes does not add significantly to the arousal induced by the presence of others. Thus, the total level of arousal may not exceed that necessary for good performance.

Any complete explanation of audience effects, however, should not ignore individual differences. People differ in their susceptibility to audience presence. Ganzer (1968), for example, with serial learning as the criterion, found that audience presence was more detrimental for high- than for low-anxious individuals. In another study Kohfeld and Weitzel (1969) concluded that susceptibility to peer group status was greater for *S*s who scored relatively low on five personality scales of the California Psychological Inventory (Ai, Achievement via Independence; Ac, Achievement via Conformance; Re, Responsibility; To, Tolerance; and Gi, Good Impression) than for individuals scoring relatively high on the scales. They argued that people who work equally hard for both low- and high-status observers tend to possess certain positive personality attributes (Ai, Ac, Re, To, Gi) to a greater degree than individuals who are more sensitive to the social status of their audience.

While most of the audience effect studies concentrated on their effects on performance, the present study focused on studying the feelings of *S*s toward the presence of others. More specifically, it investigated felt anxiety of the intrinsically oriented and the extrinsically oriented toward the presence of higher status observers with respect to two levels of task difficulty.

An interaction was expected among the three variables as indicated by the following hypotheses:

Hypothesis 1: More anxiety would be expressed by the intrinsically oriented than the extrinsically oriented while being observed performing on the *difficult* task.

Hypothesis 2: No difference in expressed anxiety would be expected between the intrinsically oriented and the extrinsically oriented while being observed performing on the *simple* task.

These hypotheses are based on two assumptions: (*a*) The intrinsically oriented emphasize achievement and task involvement more than the extrinsically oriented, and therefore, the former would be more concerned with the nature of the task than the latter. And (*b*) the presence of higher status *O*s is of concern not only to the extrinsically oriented, who stress the importance of supervision, but also to the intrinsically oriented, who might perceive *O*s as evaluators of their abilities. In view of these two assumptions, the intrinsically oriented under the observed difficult task condition are expected to experience arousal from both the task and *O*s, while arousal for the extrinsically oriented would be primarily induced by *O*s. Under the observed simple task condition, however, the main source of arousal for both groups is expected to be the presence of superior *O*s.

METHOD

Measures

The Job Attitude Scale (JAS, Saleh, 1964; Saleh & Grygier, 1969; Saleh & Lalljee, 1969) consists of 120 items involving 16 statements, each being paired with the other 15 in a forced-choice format. Six of the statements involve intrinsic factors (achievement, responsibility, recognition, advancement, nature of work, and growth in skill). The other 10 involve extrinsic factors (working conditions; company policy; salary; security; status; technical supervision; salary needs for family's sake; and interpersonal interactions with supervisor, subordinates, and equals). Job orientation was determined by scoring only the 60 items in which an intrinsic factor was paired with an extrinsic one. By giving one point whenever the intrinsic factor is checked, the possible score range is 0–60. The internal consistency (split half) of this scale was .94 and test-retest reliability was .88.

Postexperimental Questionnaire

The main item of concern in this questionnaire was the scale used by Schachter (1959) in studying the relationship between anxiety and affiliation. "Indicate how you feel about being a subject in the experiment [p. 31]." The *S*s indicated their feelings on a 6-point scale ranging from "I feel extremely uneasy" (1) to "I feel completely calm" (6). The second objective item of concern in this questionnaire involved *S*s' perceptions of the task. They were asked to indicate if the task was simple by checking one of three choices, yes, no, or ?. Scores given were 3, 1, and 2, respectively. In two open-ended questions the effects of working alone or observed were elicited.

Task

Most experiments on anxiety have used the task of rate learning of nonsense syllables. Although such a task has its experimental advantages, it was felt that the kind of learning involved is not of great importance in human life, and what there is of it seldom takes place under the kind of conditions imposed by the researcher. A better approximation to "real-life" situations is to deal with more complex intellectual activities such as problem solving. Therefore, the task chosen for this study was the solving of anagrams, a task that has been used successfully in anxiety experiments (i.e., Harleston, 1962; Russell & Sarason, 1965).

This study's anagrams were tested empirically for their difficulty by Brown (1963) on an undergraduate university sample, and they have since been used in several studies (Feather, 1965, 1966). The difficulty was defined as the percentage of the individuals who solved the anagram in 30 sec. Forty anagrams with a percentage of correct answers ranging from 51% to 80% comprised the simple task while 30 anagrams varying in difficulty from 21% to 50% formed the difficult task.

The JAS was administered to 428 first-year university male students whose participation in this experiment was a requirement of the introductory course in psychology. Scores ranged from 6 to 55. The top quarter of the sample scored 36 or more and were considered the intrinsic group (usable $N = 92$).[3] The bottom quarter scored 22 or less and they constituted the extrinsic group (usable $N = 96$; see Footnote 3). Both groups were invited to participate in the experiment that took place 4 wk. after the administration of the JAS. Each of them was further divided into four comparable subgroups (24 *S*s in each) on the basis of the JAS scores. Two intrinsic and two extrinsic subgroups worked on the simple task while the other two worked on the difficult task. In one of the subgroups each *S* worked alone while in the other, *S* was observed by two individuals who were presented as graduate students (and, hence, having higher status than *S*). Two observers were used instead of one in order to magnify the effect of being observed. The experiment, then, used a $2 \times 2 \times 2$ factorial design, two types of job orientation, intrinsic and extrinsic; two levels of task difficulty; and two working conditions, alone and observed.

The *S*s were asked to solve as many anagrams as they could in the specified time (15 min). Because it was intended that *S*s should be able to solve all anagrams in the specified time, the number solved within 15 min. was tested before the experiment on another comparable group. In the "alone" condition, *E* told *S* that he was going to leave him to work alone and that he would return after 15 min. The postexperiment questionnaire that included the dependent variables was given immediately after the experimental session.

RESULTS

The analysis of variance for the anxiety scale is presented in Table 1 which shows that the interaction effect among the three independent variables, orientation, task, and presence or absence of observers was significant ($p < .05$). Figure 1 and Table 2 show the nature of this interaction. They indicate that the intrinsically oriented *S*s expressed more anxiety than the extrinsically oriented under the observed *difficult* task condition and thus support the first hypothesis. The results also show no difference between the two groups under the observed *simple* task condition and therefore support the second hypothesis. By examining Table 2 and Figure 1 we also find no differences between the intrinsically oriented and the extrinsically oriented either under the alone simple task condition or under the alone difficult task condition.

It should be noted that Table 1 indicates two significant main effects, the task and the audience variables. In the case of the simple task, the mean anxiety score was 4.99 ($SD = .79$), and for the difficult task the mean was 4.58 ($SD = 1.12$). (The higher the mean the lower the anxiety). The mean for the alone condition was 4.94 ($SD = .95$) while it was 4.63 ($SD = 1.00$) in the observed condition.

TABLE 1

ANALYSIS OF VARIANCE OF ANXIETY SCORES AS A FUNCTION OF JOB ORIENTATION, AUDIENCE PRESENCE, AND TASK DIFFICULTY

Source	*MS*	*df*	*F*
Total	.98	189	
Between	3.24	7	
Job orientation (A)	1.84	1	2.05
Task difficulty (B)	7.79	1	8.68**
Audience presence (C)	4.40	1	4.90*
A × B	1.53	1	1.70
A × C	2.70	1	3.01
B × C	.05	1	.05
A × B × C	4.39	1	4.89*
Within	.90	182	

* $p < .05$.
** $p < .01$.

---

[3] Few *S*s did not show up for the experiment and some did not have all required data.

TABLE 2

RATED ANXIETY OF THE INTRINSICALLY ORIENTED AND THE EXTRINSICALLY ORIENTED UNDER THE AUDIENCE PRESENCE AND TASK DIFFICULTY CONDITIONS

Condition	Intrinsically oriented				Extrinsically oriented			
	Simple task		Difficult task		Simple task		Difficult task	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Alone	5.08	.86	4.83	.75	5.17	.69	4.67	1.11
Observed	4.88	.93	3.96	1.44	4.83	1.31	4.88	.83

Note.—The higher the score, the less the anxiety.

TABLE 3

ANALYSIS OF VARIANCE OF PERCEIVED TASK SIMPLICITY AS A FUNCTION OF JOB ORIENTATION, AUDIENCE PRESENCE, AND TASK DIFFICULTY

Source	*MS*	*df*	*F*
Total	.34	190	
Between	.99	7	
Job orientation (A)	.70	1	2.2
Task difficulty (B)	4.79	1	15.16**
Audience presence (C)	.07	1	.21
A × B	.30	1	.95
A × C	1.07	1	3.4*
B × C	.00	1	0.0
A × B × C	.01	1	.04
Within	.32	183	

* $p < .10$.
** $p < .001$.

These results support the assumption introduced earlier that both task difficulty and the presence of *O*s would be sources of anxiety and, therefore, arousal.

Although no specific hypothesis was introduced with regard to *S*s' perceptions of task simplicity, the results (see Table 3) tended to be consistent with those of anxiety. The main effect for the task variable was significant at the .001 level. The mean for the simple anagrams was 1.44 ($SD = .68$) and the mean for the difficult anagrams was 1.12 ($SD = .36$). (The higher the score the simpler the task.)

Table 3 also shows that the interaction effect between the orientation variable and the audience variable approached significance ($p < .10$). The nature of this interaction is shown in Figure 2, which indicates that the main difference between the intrinsically oriented and the extrinsically oriented in their perception of task difficulty was in the alone condition. In this case the first group perceived the task as being simpler ($M = 1.44$, $SD = .71$) than the second group did ($M = 1.17$, $SD = .41$) ($t = 2.25$, $p < .05$). There was no difference between the two groups under the observed condition (M of intrinsically oriented $= 1.25$, $SD = .56$; M of extrinsically oriented $= 1.28$, $SD = .57$). The results also showed a tendency on the part of the intrinsically oriented to perceive the task as being more difficult when observed than when they were alone, while a slightly reversed tendency appeared for the extrinsically oriented.

The analysis of the two open-ended questions showed reactions consistent with the previous results. For instance, the disturbing effect of *E*s' presence was more pronounced for both the intrinsically oriented and the

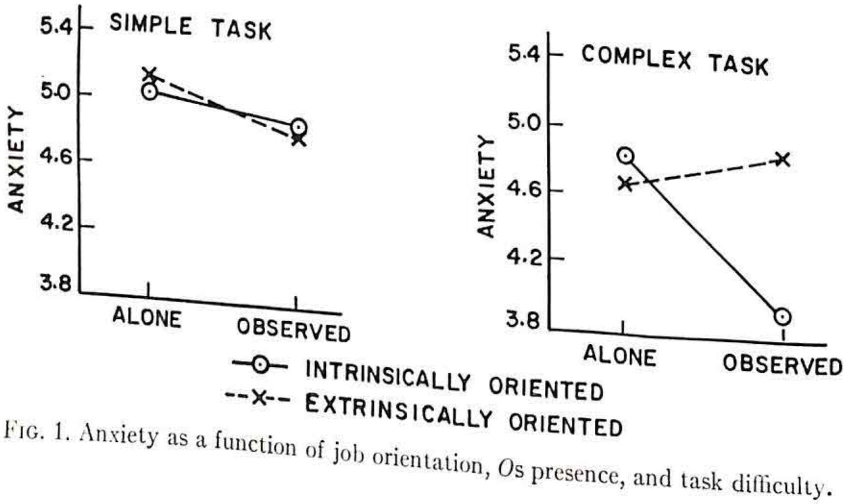


FIG. 1. Anxiety as a function of job orientation, *O*s presence, and task difficulty.

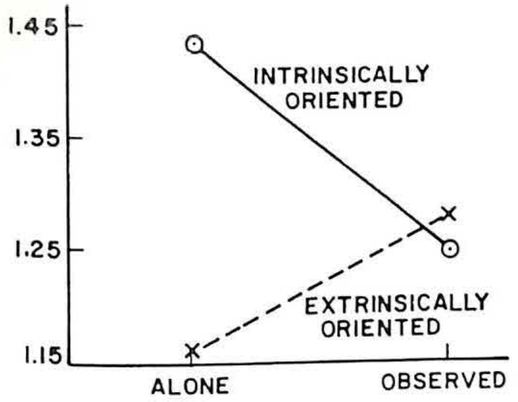


FIG. 2. Perception of task simplicity as a function of job orientation and presence of *O*s.

extrinsically oriented while working on the difficult task rather than the simple task. Furthermore, the analysis showed that *O*s' presence tended to arouse debilitating anxiety in more intrinsically oriented than in extrinsically oriented *S*s.

It is of interest to note that four *S*s in each of the observed extrinsically oriented groups mentioned that the presence of *E*s had a positive effect and improved performance. In the alone condition three extrinsically oriented *S*s stated that they would have worked better if *E* had stayed. Under no condition did any intrinsically oriented *S* mention such an effect.

DISCUSSION

The differences between rated anxiety levels of the intrinsically oriented *S*s and the extrinsically oriented *S*s in the present study could be explained in view of the differences in their job attitudes. As mentioned previously, the intrinsically oriented emphasize the task itself and the challenge it offers to test skills and abilities more than the extrinsically oriented. In a working situation the former group also would have a stronger need for achievement than the latter group. In this view, it might be argued that the intrinsically oriented more than the extrinsically oriented should be expected to concentrate on task performance and ignore the presence of *O*s. We should not expect, however, that the intrinsically oriented, who would like to be recognized for their abilities and their achievements, would not react to the presence of two higher status *O*s who watched each of them closely while he performed. Based on this understanding, the experimental situation presented this group with two sources of arousal, the task and *O*s, while the extrinsically oriented were presented only with *O*s' source. This should explain why the main difference in anxiety between the two groups was under the observed difficult task condition (see Fig. 1).

Based on the above discussion, it can be assumed that the intrinsically oriented perceived solving the anagrams not only as getting a credit toward the fulfillment of a course but also as a challenge. For them, not achieving and not performing up to a high standard would be regarded as highly undesirable, as threatening to the self-concept, and therefore would develop "achievement anxiety." It is expected that this threat and anxiety would be greater if the task was perceived more difficult. Therefore, since all *S*s perceived a different difficulty level for the two tasks of the study regardless of being observed or not (see Table 3), and since the presence of the two *O*s increased the perception of task difficulty only for the intrinsically oriented (see Fig. 2), the difficult anagrams under the observed condition would be the most difficult task for this group and consequently the most anxiety arousing.

To the extrinsically oriented, on the other hand, who are assumed to be basically less interested in successful performance, task difficulty should not be of as much concern as it is to the intrinsically oriented. Moreover, it appears that the presence of observers had some differential effect on the two groups with more positive influence on the extrinsically oriented, for whom good supervision is important.

Although the present study has presented meaningful and consistent results, the limitations of using only a single item anxiety-reaction scale and a single item task-difficulty scale should be kept in mind. Moreover, it should be pointed out that the presence of two *O*s, as in this study, might have different reactions than the presence of one *O*.

In conclusion, the different reactions of intrinsically oriented individuals and extrinsically oriented individuals to the high-status *O*s suggest implications resulting from close supervision for the two groups, particularly when the task is comprised of difficult cognitive problems. Such reactions and implications should be tested in real working situations to find out if the theory of "job orientation" presented in this laboratory study could be extended to the field.

REFERENCES

BROWN, M. Factors determining expectancy of success and reactions to success and failure. Unpublished manuscript, University of Michigan, 1963.

CUMMINGS, L. L., & EL-SALMI, A. M. Empirical research on the bases and correlates of managerial motivation: A review of the literature. *Psychological Bulletin*, 1968, **70**, 127–144.

FEATHER, N. T. Performance at a difficult task in relation to initial expectation of success, test anxiety, and need achievement. *Journal of Personality*, 1965, **33**, 206–217.

FEATHER, N. T. Effects of prior success and failure on expectations of success and subsequent performance. *Journal of Personality and Social Psychology*, 1966, 3, 287–298.

GRANZER, V. J. Effects of audience presence and test anxiety on learning and retention in a serial learning situation. *Journal of Personality and Social Psychology*, 1968, **8**, 194–199.

HARLESTON, B. W. Test anxiety and performance in problem-solving situations. *Journal of Personality*, 1962, **30**, 557–573.

HEBB, D. O. Drives and the C. N. S. (Conceptual nervous system). *Psychological Review*, 1955, **62**, 243–254.

HERZBERG, F. *Work and the nature of man*. New York: World Publishing Co., 1966.

HESS, E. H., & POLT, J. M. Pupil size in relation to mental activity during simple problem solving. *Science*, 1964, **143**, 1190–1192.

HOUSE, R. J. WIGDOR, L. A. Herzberg's dual-factor theory of job satisfaction and motivation: A review of the evidence and a criticism. *Personnel Psychology*, 1967, **20**, 369–390.

KOHFELD, D. L., WEITZEL, W. Some relations between personality factors and social facilitation. *Journal of Experimental Research in Personality*, 1969, **3**, 287–292.

RUSSELL, D. G., & SARASON, I. G. Test anxiety, sex, and experimental conditions in relation to anagram solution. *Journal of Personality and Social Psychology*, 1965, **1**, 493–496.

SALEH, S. D. A study of attitude change in the preretirement period. *Journal of Applied Psychology*, 1964, **48**, 310–312

SALEH, S. D., & GRYGIER, T. The psychodynamics of intrinsic and extrinsic job orientation. *Journal of Applied Psychology*, 1969, **53**, 446–450.

SALEH, S. D., LALLJEE, M. Sex and job orientation. *Personnel Psychology*, 1969, **22**, 465–471.

SCHACHTER, S. *The psychology of affiliation: Experimental studies of the sources of gregariousness*. Stanford: Stanford University Press, 1959.

SCHNORE, M. M. Individual patterns of physiological activity as a function of task differences and degree of arousal. *Journal of Experimental Psychology*, 1959, **58**, 117–128.

WEISSENBERG, P., & GRUENFELD, L. W. Relationship between job satisfaction and job involvement. *Journal of Applied Psychology*, 1968, **52**, 469–473.

WHITSETT, D. A., & WINSLOW, E. K. An analysis of studies critical of the motivator-hygiene theory. *Personnel Psychology*, 1967, **20**, 391–415.

ZAJONC, R. B. Social facilitation. *Science*, 1965, **149**, 269–274.

(Received August 3, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 549–553

ERROR AND VARIABILITY IN THE VISUAL PROCESSING OF DENTAL RADIOGRAPHS[1]

IRWIN L. GOLDSTEIN[2] AND WILLIAM H. MOBLEY

University of Maryland

The primary objective of the present study was to assess the magnitude of observer error and disagreement in a sample of senior dental students. The *S*s were presented with a series of 40 radiographic stimuli and were required to give a detection response, a recognition response, and a difficulty rating. Analyses of the detection responses revealed a high false positive rate and moderate miss rate. Analyses of the recognition responses showed that *S*s were fairly accurate in describing abnormalities that were correctly detected. No consistent relationship was found between observers' difficulty ratings and either detection or recognition performance.

As psychologists examine the problems surrounding the processing of complex displays (Goldstein & Allen, 1971; Howell, Johnston, & Goldstein, 1966; Kibler, 1965), it has become apparent that there are many variables and tasks that have not been subjected to experimental analysis. While psychologists have discovered that they do not have a data base governing absolute levels of performance, professionals from other fields like dentistry and medicine have begun to seriously question their ability to make judgments from their complex displays, for example, radiographic materials.

In relation to radiographic displays, Fletcher (1954) stated that,

> observer error is a subject whose importance should be more widely recognized. Until clinicians freely recognize the bluntness of the instruments they use, they will burden their students unnecessarily, deceive themselves, waste a lot of time, and reach false conclusions from which their patients may suffer unnecessarily [p. 212].

A number of studies illustrate that observer error is indeed an important problem. Garland (1950) reported that 399 chest X-rays read twice by each of three readers resulted in 32% underreading (misses) and 2% overreading (false positives). Other researchers (Yerushalmy, Harkness, Cope, & Kennedy, 1950) investigated the detection of shadows in mass chest roentgenograms and found that as many as 30% of the roentgenograms showing evidence of disease were missed and that 2% were reported as false positives.

Unfortunately, much of the medical and dental literature can be criticized because of the lack of proper experimental control. There are too few studies that present reliability data, adequately specify the sample employed, and present valid criteria for judgments. Some of the investigations (Galagan & Vermillion, 1956) used the majority opinion of the subjects in the study as a basis for deciding if pathology was present in the radiograph.

Specifically, the present study is designed to determine whether there are visual errors associated with the observation of dental radiographs. If observation errors are found, further indicators of performance like the types of errors and the confidence of the observer will also be examined.

METHOD

Subjects

Forty volunteer, male, senior dental students from the University of Maryland School of Dentistry served as *S*s. All *S*s had completed their course work in radiology and oral diagnosis, had clinical experience, and would earn their degrees within 6 mo. Each *S* received $6.00 for participating in the study.

Stimuli

The stimuli were 40 bitewing dental radiographs that had been converted to Kodachrome A slides. The film sample, selected by the Head of Oral Diagnosis at the

---

[1] This research was supported by Contract No. PH52-112 from the Education Research Program, Division of Dental Health. Appreciation is expressed to S. John Chellemi of the Division of Oral Diagnosis, University of Maryland Dental School and Morris Roseman, Division of Community Dentistry, University of Maryland Dental School for their aid in arranging for *S*s and for the many hours they spent in explaining technical details that helped establish communication between psychology and dentistry.

[2] Requests for reprints should be sent to Irwin L Goldstein, Department of Psychology, University of Maryland, College Park, Maryland 20742.

University of Maryland Dental School, was intended to be representative of films a practicing dentist might be required to diagnose. The sample included 22 positives, 15 negatives, and three slides with known processing defects.

Procedure

The films were presented via slide projector. The series of 40 stimuli were presented twice, each presentation independently randomized. For each film, *S*s were required to respond in one of three categories: (*a*) abnormality present, (*b*) processing defect present, (*c*) normal. This represented the "detection" response. In the case of the first two response categories, *S*s were asked to identify the abnormality or processing defect. This represented the "recognition" process response. Finally, *S*s were required to rate each slide on a 7-point bipolar verbally anchored scale of interpretive difficulty. Sixty seconds were allowed for viewing and responding to each slide. Following the presentation, *S*s were required to complete a postexperimental questionnaire, a self-rating of effort, a question regarding preference to avoid misses or false positives, identification of position in the experimental room, and name.

RESULTS AND DISCUSSION

Detection Error

Performance on the detection task was analyzed in terms of correct detection responses, misses, and false positives. A miss resulted when a positive film or a film with a processing defect was called "abnormality present." A detailed summary of detection performance, subdivided by presentation, is given in Table 1. For both presentations combined, that is, 80 slides, 73% of the films were correctly classified. There was a slight but nonsignificant improvement in detection performance on the second presentation.

TABLE 1

SUMMARY OF DETECTION PERFORMANCE: NUMBER AND PERCENTAGE OF CORRECT DETECTIONS, MISSES, AND FALSE POSITIVES

Performance	$\bar{X}$	%	*SD*
No. correct, both presentations	58.03	73	3.68
No. correct, Presentation 1	28.93	72	2.31
No. correct, Presentation 2	29.10	73	2.20
No. misses, Presentation 1	3.76	15	2.02
No. misses, Presentation 2	3.60	14	2.02
No. false positives, Presentation 1	7.30	41	2.39
No. false positives, Presentation 2	7.30	41	2.28

The miss rate was computed by dividing the number of misses by 25, the number of positive and processing defect films. The percentage missed for both presentations combined was 15. The false positive rate was computed by dividing the number of false positives by 18, the number of negative and processing defect films. The percentage of false positives for both presentations combined was 41.

The analysis of detection error revealed a moderate miss rate and an extremely high false positive rate. The fact that the false positive rate was higher than the miss rate is consistent with the frequent observation that students tend to overread films. It is also consistent with the indication by 75% of *S*s that they would prefer to avoid missing an abnormality as opposed to avoiding a false positive when confronted with a difficult interpretation.

The fact that the false positive rate was higher than the miss rate is also consistent with the results of Galagan and Vermillion (1956). However, a number of authors (Garland, 1950; Webber, Benton, & Ryge, 1967; Yerushalmy et al., 1950) have reported higher miss than false positive rates. Direct comparisons among studies are difficult due to differences in observers (students versus experienced radiologists and dentists) and differences in stimuli (chest X-rays versus dental film, difficulty level, and complexity of films) and differences in criteria.

Recognition and Overall Performance

The evaluation of recognition performance was accomplished by noting the correct and incorrect descriptions for positive and processing defect films that were properly classified in the detection response. On the first presentation of the film series, 10% of the positive and processing films that were correctly detected were incorrectly described. On the second presentation, 9% were incorrectly described.

These values indicate that if the observer detected the abnormality, he was fairly accurate in describing it. These data may also indicate that if the observer cannot describe what he detects, he will not report a detection. Galanter (1962) has noted that recognition error often may not be too large since detection itself may be dependent upon the observer's ability to recognize the stimulus.

In order to assess overall performance, a total score was computed for each observer. This score reflected the number of correct detection responses for negative films and the number of correct recognition responses for positive and processing defect films that were correctly detected. The results of this analysis are presented in Table 2.

The results presented in Table 2 show a slight but nonsignificant improvement in overall performance from Presentation 1 to Presentation 2. Considering both presentations combined, 33% of the films are incorrectly detected or described.

The magnitude of this error figure gives strong support to Galagan and Vermillion's (1956) proposition that diagnosis of dental radiographs is subject to considerable error, and further research effort must be directed toward understanding and reducing this error.

Reader Agreement

It is important to evaluate reader consistency as well as reader error when investigating observer performance. Intrareader agreement refers to a measure of how often single readers agree with themselves on two readings of a set of films. Interreader agreement refers to a measure of how often different readers agree among themselves over all possible combinations of readers. If any measurement is presented, pairs in agreement is the index of reader agreement most frequently employed in the radiology literature (Yerushalmy, 1955).

Intrareader agreement for the detection response was analyzed in terms of the number of first- and second-presentation responses in agreement. The mean number of pairs of detection responses in agreement was 36 or 91%.

TABLE 2

SUMMARY OF OVERALL PERFORMANCE

Performance	$\bar{X}$	%	*SD*
No. correct, both presentations	53.89	67	4.69
No. correct, Presentation 1	26.73	67	2.58
No. correct, Presentation 2	27.15	68	2.50

TABLE 3

SUMMARY OF INTRA- AND INTERREADER AGREEMENT: PROPORTION OF PAIRS IN AGREEMENT

Agreement	Detection responses	Overall performance
Intrareader agreement	.91	.90
Interreader agreement		
Presentation 1	.70	.69
Presentation 2	.70	.71

Average intrareader agreement for overall performance was 36 pairs of responses in agreement or 90%.

Interreader agreement was evaluated by noting the number of pairs of detection responses in agreement among the 40 readers over all slides. For the detection response, average interreader agreement for both presentations combined was 70%. For overall performance, average interreader agreement was also 70%. A summary of reader agreement is given in Table 3.

Since the film series was presented twice within 2 hr., the intrareader disagreement is probably an underestimate due to short-term recall. However, even within this relatively short time span, the intrareader disagreement was 9% for the detection responses and 10% for overall performance. These figures are comparable to the 21% intrareader disagreement reported in the chest radiology literature by Yerushalmy et al. (1951) and the 6% reported by Galagan and Vermillion (1956) for dental radiographs. The 30% interreader disagreement for detection responses and 30% for overall performance found in the present study are comparable to the 30% reported by Galagan and Vermillion (1956).

The relatively high intrareader disagreement indicates that *S*s were less than completely consistent in their own interpretations even within the short retest period. The relatively high interreader disagreement suggests that *S*s had not employed a common frame of reference or stereotype in interpreting the films, exhibited differential response bias, and/or differed in abilities contributing to performance on this task.

Another measure of reliability, the test-retest correlation, was obtained in order to have information about the stability of the order of individuals over time. In the present study, the retest was within 2 hr. and, thus, is biased due to short-term memory. However, even within this short period, the test-retest correlation of .56 for detection is quite low.

Further analyses showed that there were either no errors or a very small number of errors on over one-third of the films and that the low or no error films were generally positive. To the extent that the film series was a representative sample, there appears to be a relatively large group of positive films that offer the observer no difficulty in detecting the existence of an abnormality. On the other hand, the poor performance on the remaining slides is even more striking. The problem for further research is to isolate the factors that discriminate between these easy detections and difficult detections. The increase in the test-retest reliability to .72 for overall performance may in part be attributed to the increased variance in total score and to the fact that overall performance includes a broader sample of observer behavior.

In summarizing the reader agreement and reliability results it may be concluded that, under the conditions of the present study, interpretation of dental radiographs is not a highly stable, consistent process. Further research is needed to establish whether the relatively low test-retest reliability was a function of the particular films, *S*s, and conditions of the present study. In view of the similarity of the reader agreement results in this study and earlier studies, it may well be that the low reliability encountered was not specific to this study but may be a more pervasive problem.

Additional Analyses

The relation between difficulty rating and performance (right-wrong) was analyzed by computing point biserial correlation coefficients. A coefficient was computed for each *S* to reduce measurement error attributable to individual differences in interpretation of the difficulty rating scale. The t ratio was significant at the $p < .05$ level for only 17 of the 40 *S*s with a mean r_{pb} of .23. For the detection responses, significance was at the $p < .05$ level for only 13 of the 40 *S*s with the average coefficient being .19.

The reliability of the difficulty ratings was assessed using the procedure outlined by Ebel (1951) and Guilford (1954, 1965). This method provides an estimate of the reliability of the composite ratings for an entire group and an estimate of the reliability of a single rater. For Presentation 1, the estimated composite reliability was .87 and the estimated reliability for a single rater was .14. For the Presentation 2 these values were .90 and .18, respectively. Since the primary interest was in the single observer, the .14 and .18 reliability estimates were of particular interest. These results indicate that there was very little relationship between different observer's ratings of the same films.

The relations between performance and the self-ratings of performance and effort were examined by computing product-moment correlations between overall performance score and each of the two self-ratings. The correlation between total score and the self-rating of performance was .08, while the correlation between total score and the self-rating of effort was −.09. Neither coefficient was significantly different from zero.

The fact that the relation between difficulty rating and item performance was significant for only 17 of 40 *S*s in the case of overall performance and 13 of 40 *S*s in the case of detection performance suggests that there is, at best, a weak relationship between these two variables. This finding, along with the low reliability of an observer's difficulty ratings and the nonsignificant correlation between self-rating of performance and total score indicates that *S*s were not able to adequately assess difficulty and their own performance. The findings are consistent with Yerushalmy et al.'s (1955) failure to find a significant relationship between difficulty and performance.

These results are of significance to those who recommend dual reading of "difficult" radiographs. If a reader is unable to identify which radiographs are difficult, the radiographs he passes on for dual reading may not be the ones he is in fact misreading. The above results, combined with the relatively high reader disagreement and low reliability, suggest that dual reading is not a panacea for observer error.

Limitations of Generalizations

It is important to note that the students were not permitted to clinically examine the patient, nor did they have access to the patient's history. Also, senior dental students rather than dentists or radiologists were used as *S*s. While it may be assumed that *S*s were trained observers, it is apparent that they differ from practicing dentists in experience, age, set, and other variables that may be relevant to performance on this task. There are also details about the experimental procedures that limit generalizability. In this study, the students read the X-rays as presented by a slide projector rather than viewing the film against a light source. Even though the students were familiar with the technique because similar methods were used in their classroom work in dental radiograph interpretation, this is not the normal procedure used by practicing dentists. Also, the time allowance of 60 sec. that was judged as completely adequate by all *S*s did place a time limitation on the task, which is not common to normal dental interpretation tasks. Finally, as with any study, generalizability is restricted by the particular films and methods used.

Future research with different stimuli and observers is necessary in order to examine the issues raised in this experiment. Particular attention should be given to the high false positive rate that has been found in several different investigations (Galagan & Vermillion, 1956). Future research may find it fruitful to focus on the perceived rewards and costs associated with the available response alternatives.

REFERENCES

EBEL, R. L. Estimation of the reliability of ratings. *Psychometrika*, 1951, **16**, 407–424.

FLETCHER, C. M. Observer error. *Lancet*, 1954, **1**, 212.

GALAGAN, D., & VERMILLION, J. Diagnosis of caries by radiographic interpretation. *Journal of Dental Research*, 1956, **35**, 33–38.

GALANTER, E. Contemporary psychophysics. In R. Brown (Ed.), *New directions in psychology*. Vol. 1. New York: Holt, Rinehart, 1962.

GARLAND, L. H. On the reliability of Roentgen survey procedures. *American Journal of Roentgenology*, 1950, **64**, 32–41.

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1965.

GOLDSTEIN, I. L., & ALLEN, J. C. Effects of irrelevant stimuli on the processing of information in complex displays. *Journal of Applied Psychology*, 1971, **55**, 110–113.

HOWELL, W. C., JOHNSTON, W. A., & GOLDSTEIN, I. L. Complex monitoring and its relation to the classical problems of vigilance. *Journal of Organizational Behavior and Human Performance*, 1966, **1**, 129–150.

KIBLER, A. W. The relevance of vigilance research to aerospace monitoring tasks. *Human Factors*, 1965, **7**, 91–99.

WEBBER, R., BENTON, P., & RYGE, G. Diagnostic variation in radiographs. Paper presented at the meeting of the International Association for Dental Research, Washington, D. C., March, 1967.

YERUSHALMY, J. Reliability of chest radiography in the diagnosis of pulmonary lesions. *American Journal of Surgery*, 1955, **89**, 231–240.

YERUSHALMY, J., GARLAND, L. H., HARKNESS, J. T., HINSHAW, H. C., MILLER, E. R. SHIPSMAN, S., & ZERLING, H. G. An evaluation of the role of series chest Roentgenograms in estimating the progress of disease in patients with pulmonary tuberculosis. *American Review of Tuberculosis*, 1951, **64**, 225–248.

YERUSHALMY, J., HARKNESS, J. F., COPE, J. H., & KENNEDY, B. R. The role of dual reading in mass radiology. *American Review of Tuberculosis*, 1950, **61**, 443–464.

(Received July 23, 1970)

ournal of Applied Psychology
971, Vol. 55, No. 6, 554–558

SPATIAL AND TEMPORAL INFORMATION AS CUES IN A TIME-SHARING TASK[1]

PETER W. DORFMAN AND IRWIN L. GOLDSTEIN

University of Maryland

The effects of directional and nondirectional cues were investigated on a task where *S*s responded to moving visual stimuli that entered critical zones in each of three visual displays. The *S*s given directional cues that supplied spatial and temporal information performed significantly better than *S*s in the nondirectional group, who received only temporal cues, and a control group that did not receive any experimental cues. Furthermore, *S*s in the control group performed as well as *S*s given nondirectional cues. Thus, temporal cues did not lead to enhanced performance unless the display requiring attention was designated by a spatial cue. These data also indicated that increased speed stress led to linear decrements in performance.

Today's complex man-machine systems frequently include human tasks that require the operator to attend and respond to multiple-information sources. Weisz and McElroy (1964) and others (Allen, 1969; Jeantheau, 1959; Mackworth & Mackworth, 1958; McCormick, 1964; Mowbray & Gebhard, 1961; Olson, 1963) have found that complex operator tasks may momentarily overload the information processing capacity of the individual resulting in the deterioration of performance. These data have led to the common-sense notion that equipment design must specify the acceptable rate of input information (McCormick, 1970). Unfortunately, there are man-machine situations (e.g., air traffic control) where speed stress is an essential aspect of the task. In these tasks, especially those involving a perceptual-motor component, response timing has been postulated as an important factor in coping with increases in speed stress.

Timing is defined by Conrad (1955, 1956) as the characteristic of skill that creates the most favorable temporal conditions for a response. Good timing produces a temporal layout of a series of responses made at the most appropriate moment and leading to smooth performance while bad timing results in irregular, jerky performance. Schmidt (1968) has postulated that high levels of skill attained by well-practiced *S*s would not be possible without proper timing. Timing difficulties have been noted by monitors who must attend to several displays. In many of these situations, the rates of change on each display are different, and the monitor has to make critical decisions based on the moment that the various pointers have reached their prescribed levels. McCormick (1970) specifically notes that monitors in time-sharing tasks should be directed toward these subtasks which require immediate attention.

Cueing as a technique provides information that should permit *S*s to predict and anticipate critical stimuli that require immediate attention. While research (Mudd, 1965; Mudd & McCormick, 1960; Pollack & Teece, 1958; Smith, Farrell, & Gonzalez, 1965) investigating the effect of cueing has established its effectiveness, the relationship of cueing to performance with a timing task at different speed stress levels is not nearly so clear. Thus, the purpose of the present research was to investigate the effects of spatial and temporal cues on performance in a perceptual-motor task.

METHOD

Subjects

Seventeen male volunteer psychology students from the University of Maryland served as *S*s.

Apparatus

The apparatus consisted of a visual display, a response panel, and an operator control panel. The visual

---

[1] Computer time was supported in full through the facilities of the Computer Science Center at the University of Maryland. The research was supported by a grant to the Center for Language and Cognition from the Biomedical Science Support Committee of the University of Maryland.

[2] Requests for reprints should be sent to Irwin L. Goldstein, Department of Psychology, University of Maryland, College Park, Maryland 20742.

display was composed of three lighted dials, 4.5 in. long and 1.5 in. high, placed side by side approximately 6 in. apart. Two black vertical target lines marked off each dial into three areas or zones. The pointers, one on each of the three dials, moved at a constant velocity to the right and then to the left. The pointer stopped when it reached the end of the dial if *S* failed to push the response button and thereby reverse the direction of the pointers. A light source mounted on the top of each of the three dials provided an anticipatory warning signal (visual cue) that was electronically controlled and became activated when the pointer approached a target line.

Response buttons mounted on a panel in front of *S* were located in spatial correspondence with the three dials and placed directly in front of *S*. The *S* placed one hand on the response panel and used his index, third, and fourth fingers to push the response buttons.

The movements of the three pointers were permanently recorded by three Esterline Angus chart recorders. There was an exact correspondence between the pointer movement and the record on the chart because each pointer was attached directly to the chart recorder pen. Thus, the following events were recorded: the location and moment at which a pointer reversed direction, the moment at which a pointer stopped and then restarted, and the time at which a signal occurred.

The *E* controlled the velocity of the pointers and the cueing conditions by means of a control panel. The *S* sat 7½ ft. from the display and was screened from *E* and the operator panel.

Task

The *S* was instructed to attend to the three dials on the visual display and reverse the direction of the pointers on each dial by pushing the response button corresponding to the pointer whenever it crossed the target lines. The target lines divided each dial into a central zone and two border zones, one on each side. The *S* was instructed to keep the pointers continuously in motion during the task; failure to reverse the direction of the pointer resulted in the pointer stopping at the end of the dial. Thus, *S* was required to determine the order of the arrival of signals (pointers crossing the target line) in the three different dials as well as timing his response to the individual signals. Visual cues provided *S* with additional information to determine the order of the arrival of signals.

Cueing Conditions

There were three visual cueing conditions: directional, nondirectional, and no cues. For the direction-oriented (spatial and temporal) condition, the light above each dial only became activated when the pointer on that dial approached the border zone. In the nondirection-oriented (temporal) condition all of the lights were activated simultaneously when any one of the three pointers approached a border zone. The direction-oriented cues therefore contained more information than the nondirection-oriented cues, as the former cues pointed out the exact dial that needed attention while the latter informed *S* that one of the three dials needed attention. The *S*s in the control condition were not provided with any cueing information.

Speed Stress Conditions

The levels of speed stress were determined by the potential number of signals per trial where a signal was defined by a superimposition of a pointer on a target line. The total number of signals per trial consisted of the sum of the number of signals per trial per dial. The levels of speed stress were 29, 39, 53, and 73 signals per 45-sec. trial. For ease of expression, the speeds were labeled as: slow, 29 signals per trial; medium, 39 signals per trial; medium-fast, 53 signals per trial; and fast, 73 signals per trial. The velocity of a given pointer during a 45-sec. trial remained constant, but each pointer had a different velocity from the other pointers.

Procedure

Each experimental session lasted approximately 1 hr. and 20 min., and was composed of 12 practice trials (3 cue conditions by 4 speed conditions) and 24 experimental trials (2 trials for each of the three cue conditions by 4 speed conditions) each lasting 45 sec. with 30-sec. intertrial intervals. When the initial practice trials were completed and all questions were answered, *S*s were given a 10-min. rest before proceeding through the experimental trials. The order of presentation of cueing condition was randomized for each *S* as well as the order of presentation of speeds within the cueing conditions.

RESULTS

The data were analyzed by a two-factor ANOVA with repeated measurements over both factors. Tukey's multiple-comparison method was used to determine the significance of differences in mean performance in the three cueing conditions and in the four levels of speed stress.

The *S*s' responses were scored according to the number of correct reversals (hits). Hits were determined by the displacement of *S*'s response from the target line and were recorded when *S* responded within .2 cm. before the target line or within 1.1 cm. after the target line. A measure of percentage of signals correct was obtained by dividing the number of hits for the dial per trial by the potential number of hits possible per dial per trial. The percentage of correct responses in each dial was averaged with the percentage of correct responses in th other two dials to form a composite measure of performance per trial.

An analysis of variance of the data indicated that both speed and cueing conditions were

TABLE 1

PERCENTAGE OF CORRECT RESPONSES

Speed	Cueing		
	Directional	Nondirectional	Control
Slow	82	73	74
Medium	74	65	60
Medium-fast	54	45	44
Fast	35	27	27

significant (for cueing: $F = 21.02$, $df = 2, 32$, $p < .001$; for speed: $F = 457.24$, $df = 3, 48$, $p < .001$), but the interaction effect of speed and cueing was not significant ($F = 1.36$, $df = 6, 96$, $p > .05$). Table 1 shows the percentage of correct responses for all experimental conditions.

The effects of speed stress on performance are consistent, with performance deteriorating with increased speed stress. Tukey's multiple comparison indicated that each level of stress was significantly different from all other levels of stress. An examination of these data showed that performance in the directional cueing condition consistently exceeded performance in the nondirectional and control conditions for all levels of speed stress. Performance in the nondirectional condition was only slightly superior to the control condition in the medium and medium-fast speeds and slightly inferior or equal to the control condition in the slow and fast speeds. The results of a Tukey analysis indicated that the directional-cueing condition was significantly different from the nondirectional ($p < .01$) and control-cueing conditions ($p < .01$). The nondirectional condition was not significantly different from the control condition ($p > .05$). Thus, performance for the directional-cueing conditions exceeded performance in the control and nondirectional cueing conditions, while the latter two were essentially equal.

DISCUSSION

Speed Stress

As might be expected, many researchers (Alluisi, Muller, & Fitts, 1957; Conrad, 1951; Gould & Schaffer, 1967; Jeantheau, 1959; Mackworth & Mackworth, 1956, 1958; Nicholson, 1962; Olson, 1963; Wagner, Fitts, & Noble, 1954) have found performance decrements with increased speed stress. (A ω^2 analysis for the present study indicated that speed stress accounted for 67% of the variance.) The form of the relationship, however, is not nearly as well documented. Conrad (1951), warning that utmost care should be used when comparing the speed-performance relationship in different tasks, found a logarithmic relationship between speed and performance using errors of omission as the dependent variable. Using errors of timing as the criterion of performance, however, Conrad found that performance was not affected by speed stress. In the present study, when the four speed conditions are examined according to the potential hits possible for each speed stress condition, performance retained an essentially linear relationship with increased speed stress resulting in decreased performance. Thus, the specific form of the relationship between speed stress and performance depends upon the particular task and the criterion of performance.

Cueing

The analysis of variance had indicated that the cueing variable was significant at the $p < .001$ level, but only accounted for 4% of the variance as determined by the ω^2. The following two explanations could account for the above finding: (*a*) All *S*s performed significantly better in the directional-cueing condition, but the magnitude of increased performance was small. Or (*b*) most *S*s performed equally well with or without cueing, but a few *S*s performed exceptionally well. When percentage correct reversals are summed across speed conditions, all *S*s performed better in the directional condition than in the control condition, and 15 of 17 *S*s responded better in the directional condition than in the nondirectional condition. Thus, the reason for the small amount of variance accounted for by the cueing variable is that *S*s consistently responded better in the directional condition, but the magnitude of difference was small.

In order to further examine individual consistency, Spearman-Brown composite reliabilities were computed using the individual reliabilities obtained by correlating *S*'s score on each of the cueing conditions with scores on the other cueing conditions as well as *S*'s scores in each speed condition with every other speed condition. The reliabilities were .93 for the cueing condition and .98 for the speed condition. Thus, *S*s' relative level of performance in one treatment condition predicted their level of performance relative to other *S*s in the other treatment conditions. These results which indicated individual differences in performance remained stable across both speed and cueing conditions.

In conclusion, *S*s consistently performed better in the directional cueing condition where there was a spatial as well as a temporal cue. It appears that in this task where there was a time-sharing (three visual displays) and a speed stress dimension, a temporal cue, by itself, did not provide much additional information beyond what the control group obtained from simply observing the display. In some cases, the design engineer has not considered bringing *S*s' attention to the display by using a spatial cue or some association between a particular cue and a particular display. While it is difficult to generalize from visual to auditory cues, it is interesting to note that there has been a growing concern over the use of an auditory cue if it just supplies temporal information. The difficulties related to associating such a cue with a particular display may, in part, furnish an explanation for the pilot's skepticism of "a growing cacophony of zoo noises [Hunter, 1969, p. 71]" that are being used as warning signals in aircraft.

One interesting question about the utility of a temporal cue relates to the a priori probability of a specific display requiring attention. It is possible that temporal cues will suffice in those conditions where *S*s have learned or have been informed that particular displays are more likely to necessitate attention during certain temporal intervals. Further research is planned to investigate this issue.

REFERENCES

ALLEN, J. C. The use of tactics in a psychomotor skill. Unpublished doctoral dissertation, University of Maryland, 1969.

ALLUISI, E. A., MULLER, P. F., & FITTS, P. M. An information analysis of verbal and motor responses in a forced-paced serial task. *Journal of Experimental Psychology*, 1957, **53**, 153–158.

CONRAD, R. Speed and load stress in a sensory-motor skill. *British Journal of Industrial Medicine*, 1951, **8**, 1–7.

CONRAD, R. Some effects on performance of changes in perceptual load. *Journal of Experimental Psychology*, 1955, **49**, 313–322.

CONRAD, R. The timing of signals in skill. *Journal of Experimental Psychology*, 1956, **51**, 365–370.

GOULD, J. D., & SCHAFFER, A. The effects of divided attention on visual monitoring of multi-channel displays. *Human Factors*, 1967, **9**, 191–202.

HUNTER, G. Interest rising in voice warnings. *Aviation Week and Space Technology*, March 3, 1969, 68–74.

JEANTHEAU, G. *The differential effects of speed and load stress on task performance.* (Contract AF 33(616) 3612, Proj. 7184, Task 71583, WADC TR 59 7) Wright-Patterson Air Force Base, Ohio: *United States Air Force Aero Medical Lab*, 1959.

MACKWORTH, J. F., & MACKWORTH, N. H. The overlapping of signals for decisions. *American Journal of Psychology*, 1956, **69**, 26–47.

MACKWORTH, N. H., & MACKWORTH, J. F. Visual search for successive decisions. *British Journal of Psychology*, 1958, **49**, 210–221.

MCCORMICK, E. J. *Human factors engineering.* (2nd ed.) New York: McGraw-Hill, 1964.

MCCORMICK, E. J. *Human factors engineering.* (3rd ed.) New York: McGraw-Hill, 1970.

MOWBRAY, G. H., & GEBHARD, J. W. Man's senses as informational channels. In H. W. Sinaiko (Ed.), *Selected papers on human factors in the design and the use of control systems.* New York: Dover, 1961.

MUDD, S. A. Experimental evaluation of binary pure tone and auditory displays. *Journal of Applied Psychology*, 1965, **49**, 112–121.

MUDD, S. A., & MCCORMICK, E. J. The use of auditory cues in a visual search task. *Journal of Applied Psychology*, 1960, **44**, 184–188.

NICHOLSON, R. M. Maximum information-handling rates for sequentially presented visual stimuli. *Human Factors*, 1962, **4**, 367–373.

OLSON, P. L. Variables influencing operator information processing. *Human Factors*, 1963, **5**, 109–116.

POLLACK, I., & TEECE, J. Speech annunciator warning indicator system: Preliminary evaluation. *Journal of the Acoustical Society of America*, 1958, **30**, 58–61.

SCHMIDT, R. A. Anticipation and timing in human motor performance. *Psychological Bulletin*, 1968, **70**, 631–646.

SMITH, S., FARREL, R. J., & GONZALEZ, B. K. Effects of control-display compatibility and monitoring cues on multiple-task performance. *Perceptual Motor Skills*, 1965, **20**, 781–785.

WAGNER, R. C., FITTS, P. M., & NOBLE, M. E. *Preliminary investigation of speed and load as dimensions of psycho-motor tasks.* (TR54-45) Lackland Air Force Base, Texas: Air Force Personnel and Training Research Center, Air Research and Development Command, 1954.

WEISZ, A. Z., & MCELROY, L. S. Information processing in a complex task under speed stress. (ESD-TDR-64-391) Bedford, Massachusetts: Decision Sciences Laboratory, United States Air Force, L. G., Hanscom Field, 1964.

(Received August 28, 1970)

Manuscripts Accepted for Publication in the *Journal of Applied Psychology*

Human Performance Effectiveness and the Systems Measurement Bed: J. E. Uhlaner*: United States Army Behavioral Science Research Laboratory: Department of the Army, 1320 Wilson Boulevard, Arlington, Virginia 22209.

Use of Smallest Space Analysis in Studying Scale Structure: An Application to the California Psychological Inventory: Eliezer S. Karni,* and Joseph Levin: Department of Psychology, Tel-Aviv University, Tel-Aviv, Israel.

Training, Motivation, and Personality as Determinants of the Effectiveness of Brainstorming Groups and Individuals: Thomas J. Bouchard, Jr.*: Department of Psychology, University of Minnesota, 205 Elliott Hall, Minneapolis, Minnesota 55455.

Equivalence of Alternate Forms of Six Psychometric Measures during Repeated Testing: Joseph J. Fleishman,* and E. Ralph Dusek: Behavioral Sciences Laboratory, United States Army Research Institute of Environmental Medicine, Department of the Army, Natick, Massachusetts 01760.

Prediction of Cigarette Consumption Level with Personality and Socioeconomic Variables: Sadrudin A. Ahmed*: Division of Business Administration, Bishop's University, Lennoxville, Quebec, Canada.

Image of Political Candidates: Values and Voter Preference: Leonard V. Gordon*: School of Education, State University of New York at Albany, Room 114, Albany, New York 12203.

Relationship between Porter's Need Satisfaction Questionnaire and the Job Description Index: Nicholas Imparato*: College of Business Administration, University of San Francisco, San Francisco, California 94117.

Rating of Salespersons by Male Customers: 1971: Robert A. Peterson*: Department of Marketing Administration, Graduate School of Business, University of Texas at Austin, Austin, Texas 78712.

Estimating the Interjudge Reliability of the Ali-Frazier Fight: William M. Stallings,* and Gerald M. Gillmore: Measurement and Research Division, University of Illinois, 307 Engineering Hall, Urbana, Illinois 61801.

Training of a Perceptual Skill by Either Rewarding or Aversive Feedback Compared on Efficiency, Transfer, and Stress: Dirk C. Prather,* Gene A. Berry, and Gerald L. Jones: Department of Psychology and Leadership, Department of the Air Force, United States Air Force Academy, Colorado 08040.

Task Difficulty and the Efficiency of Muscular Activity: Stjepan Vidaček,* and Julius Wishner: Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104.

Reliability Estimated by Scoring Key Equivalence across Samples: Lawrence R. James,* Robert L. Ellison, David G. Fox, and Calvin W. Taylor: Life Insurance Agency Management Association, 170 Sigourney Street, Hartford, Connecticut 06105.

Relations of Ability and Interest with Currency of Professional Knowledge among Engineers: H. G. Kaufman*: 25 Jones Street, New York, New York 10014.

Effects of Interpersonal Group Composition on the Problem-Solving Behavior of Middle Managers: W. Brendan Reddy,* and Anne Byrnes: Department of Psychology, University of Cincinnati, Cincinnati, Ohio 45221.

Brainstorming on a "Hot" Problem: The Effects of Training and Practice on Individual and Group Performance: Peter C. Dillon, William K. Graham,* and Andrea A. Aidells: Department of Psychology, University of California at Berkeley, 3210 Tolman Hall, Berkeley, California 94720.

Climate, Economy, and the Differential Migration of Black and White Workers: Bernard M. Bass,* and Ralph A. Alexander: Management Research Center, Graduate School of Management, University of Rochester, Rochester, New York 14627.

Effect of Minimizing Coercion on the Rehabilitation of Prisoners: Douglas A. Bigelow,* and Richard H. Driscoll: Department of Psychology, University of Colorado, Boulder, Colorado 80302.

Liking and Loading: Carole Golightly, Devereaux Maitland Huffman, and Donn Byrne*: Department of Psychology, Purdue University, Lafayette, Indiana 47907.

Volunteer Satisfaction with In-Country Training for the Peace Corps: Reanalyses and Extended Findings: Richard R. Jones,* and W. J. Burns: Oregon Research Institute, 488 East Eleventh Avenue, Eugene, Oregon 97403.

Relationship between Measures of Effort and Job Performance: William E. Williams,* and Dale A. Seiler: Western Electric Company, Incorporated, Engineering Research Center, P. O. Box 900, Princeton, New Jersey 08540.

---

* Asterisk indicates author for whom address is supplied.

Journal of Applied Psychology
1971, Vol. 55, No. 6, 559-563

EFFECTS OF ALCOHOL INGESTION ON TRACKING PERFORMANCE DURING ANGULAR ACCELERATION[1]

WILLIAM E. COLLINS
DAVID J. SCHROEDER
Civil Aeromedical Institute
Federal Aviation Administration
Oklahoma City

AND

RICHARD D. GILSON
FREDERICK E. GUEDRY, JR.[2]
United States Navy Aerospace
Medical Research Laboratory
Pensacola, Florida

Following practice and baseline tests of tracking performance in both static (stationary) and dynamic (whole body angular accelearation) conditions, 10 *S*s received orange juice that contained 2.0 ml. of 100- proof vodka per kg. of *S* weight; others drank a control beverage. Tests, conducted 1–10 hr. after drinking, were in total darkness with the exception of the illuminated visual display. Static tracking errors for alcohol *S*s were significantly higher than those of controls only at the 4-hr. session. However, alcohol *S*s made significantly more dynamic tracking errors during the 1-, 2-, and 4-hr. sessions and, concomitantly, had significantly more nystagmic eye movements and higher rates of slow phase eye velocity than did controls. Although eye-hand coordination may show little or no impairment following alcohol ingestion in static situations, it may be seriously degraded during motion.

Schroeder (1971a, 1971b) has shown that the ingestion of alcohol depresses both ocular nystagmus (a pattern of alternate slow sweeps and fast return movements of the eyes) and vertigo sensations to rotatory or caloric vestibular stimulation when *S*s are in darkness. But in illumination, similarly provoked nystagmus is considerably stronger than it is normally (Heifer, 1969; Schroeder, 1971a, 1971b). This stronger-than-normal nystagmus in the light is not related to changes in vestibular sensitivity but to the depressive action of alcohol on the visual fixation mechanism. In the absence of alcohol involvement, nystagmus in illumination is ordinarily abolished or substantially inhibited by visual fixation. For example, figure skaters typically use visual fixation following an on-ice spin to suppress both nystagmus and vertigo (Collins' 1966). However, the alcohol-induced increase of ocular nystagmus occasioned by vestibular stimulation in the light suggests a decrease in the capacity to control blurring of the visual field and to inhibit dizziness. This suggestion poses some obvious questions regarding the ability of men to perform visual tasks adequately during vestibular stimulation after drinking alcohol. Most studies of the effects of alcohol on human performance involve static situations, that is, situations in which the men are not subjected to motion. It is conceivable that the addition of motion, which is involved in a variety of activities such as piloting an aircraft or driving an automobile, might produce deleterious effects on performance not usually obtained in static situations. Therefore, the present study was designed to examine some of the effects of alcohol ingestion on visual tracking performance during angular accelerations.

METHOD

Subjects

Twenty male college students, paid volunteers ranging in age from 21 to 30 yr., served as *S*s. None had previous laboratory experience involving vestibular stimulation. Each *S* was assigned at random to one of two equal groups, a control and an alcohol group.

Apparatus

Angular acceleration was supplied by a Stille-Werner RS-3 rotator fitted with a small cockpit in which *S* was enclosed and seated upright with his head anteverted slightly so that his horizontal semicircular canals were approximately in the plane of rotation. (The head was inclined so that an imaginary line between the outer canthus of the eye and the tragus of the ear were parallel to the floor of the room.) A fitted headrest helped to maintain the desired position. A triangular waveform input from a Wavetek Model 155 waveform generator was used as a command signal for the rotator. The velocity of the latter was proportional to the input voltage, and a peak angular velocity of 120° per second was attained in both the clockwise and counterclockwise directions. The waveform period was 48 sec.

---

[1] This study was conducted by the Federal Aviation Administration's Civil Aeromedical Institute and was cosponsored by the Federal Aviation Administration, the Army (United States Army Medical Research and Development Command) and the Navy (United States Navy Aerospace Medical Research Laboratory, Pensacola, Florida). The expert assistance of Carl Moore in obtaining the blood samples and of Elizabeth Gilson, Carlyn Manley, Cynthia Cochran, and Blair Fennell in the collection and scoring of data are gratefully acknowledged. Gas chromatographic analyses of the blood samples were performed under contract by Kurt Dubowski, University of Oklahoma Medical Center.

[2] Requests for reprints should be sent to William E. Collins, Psychology Laboratory AC-118, Civil Aeromedical Institute, Federal Aviation Administration, P. O. Box 25082, Oklahoma City, Oklahoma 73125.

A compensatory visual tracking task provided both a direct practical measure of performance and an indirect measure of acuity. The tracking apparatus has been described in detail elsewhere (Gilson, Guedry, & Benson, 1970). Briefly, a 14-sec. sinusoidal "forcing function" input deflected the vertical needle of an aircraft localizer/glide-slope indicator while *S* attempted to maintain the needle in the null position by manipulation of a control stick. Deviations of the needle from this position were considered errors ("absolute" error), and a voltage proportional to these deviations was electronically integrated over consecutive 1-sec. intervals throughout a trial.

The display was illuminated by a 3 v. dc bulb mounted in a tube in front of *S* but below his line of sight. Light was projected through the tube to localize on the display and to minimize reflection in the otherwise darkened room. The luminance of the display was measured with the aid of a card sprayed with the same white paint as the display needles. This card was placed in the light, just in front of the display, and measurements were made with a MacBeth illuminometer from *S*'s viewing position. The voltage across the bulb was adjusted until the luminance was 1 ft l., a level recommended for aircraft instruments (Morgan, Cook, Chapanis, & Lund, 1963, pp. 76–81).

Recording

Silver disk electrodes taped to the outer canthi of the eyes and a reference electrode on the forehead were used to record eye movements by the corneoretinal potential method with a 3-sec. preamplification time constant. Calibration of horizontal eye movements was accomplished with two small, alternately flashing lights, horizontally separated to subtend a visual angle of 15°. Integrated tracking error, eye movements, and rotational velocity were simultaneously displayed on an Offner Type T electroencephalograph.

Alcohol Ingestion

The *Ss* in the alcohol group consumed a mixture of 100-proof Smirnoff vodka and reconstituted (frozen concentrate of) orange juice. The mixture (900 ml.) contained 2.0 ml. of vodka per kg. of body weight. (For a 165-lb. *S*, this would amount to about 5 fluid oz. of 100-proof liquor.) Control *Ss* received only orange juice with a few drops of rum extract added to alter the odor and taste of the beverage. (They were led to believe that they were drinking alcohol.) All *Ss* consumed their drinks within a 30-min. period.

Procedure

Prior to being tested, each *S* was given 5 min. of tracking practice with the cockpit stationary. The experimental sequence that followed comprised six testing sessions: a predrinking session (pre) and five postdrinking sessions at 1, 2, 4, 8, and 10 hr. after completion of drinking. All sessions, both practice and experimental, were conducted with the room in total darkness with the exception of the visual display. Immediately before each testing session, venous blood samples of from 3 to 5 cc. were drawn for analysis of blood alcohol levels by gas chromatography.

Each session consisted of a 2-min. "static" tracking trial with the cockpit stationary and a 4-min. "dynamic" tracking trial with the cockpit rotating through five complete cycles (240 sec.). The order of these trials was alternated across *Ss*, and at least a 1-min. interval was allowed between trials. Eye-movement calibrations were obtained prior to each period of dynamic tracking.

Scoring

The raw tracking errors for 1-sec. intervals were summed, and an average value was obtained for each static and dynamic trial.

Measures of nystagmus included the number of nystagmic beats and the amount of slow phase eye displacement occurring in three 5-sec. intervals during each dynamic tracking trial; one sampling interval was 32–37 sec. from the start, one was 131–136 sec. from the start, and the other was 10–15 sec. from the end of each trial. The sampling intervals were chosen to include maximum output of nystagmus in a single direction near the beginning, middle, and end of each test period. Mean values in degrees per second or beats per second were calculated and used to represent eye-movement activity.

RESULTS

The following average values of alcohol (ethanol) were obtained from the alcohol group for the pre, 1-, 2-, 4-, 8-, and 10-hr. blood samples, respectively: 0%, .074%, .073%, .047%, .001%, and 0%. (Note that the lowest value specified in state motor vehicle statutes for presumption that a driver is under the influence of liquor is .080% [Erwin, 1970].) Control group *Ss* yielded no evidence of ethanol in their samples.

Means and standard deviations for the slow phase and frequency measures of nystagmus and for the tracking error in both static and dynamic conditions appear in Table 1. Changes in performance across sessions are shown in Figure 1 where they are presented as per-

TABLE 1

MEANS AND STANDARD DEVIATIONS BY SESSION FOR TRACKING ERROR (ARBITRARY UNITS), SLOW PHASE NYSTAGMUS (DEGREES PER SECOND), AND FREQUENCY OF NYSTAGMUS (BEATS PER SEC)

Measure	Condition	Group	Mea-sure	Session					
				Pre	1 hr.	2 hr.	4 hr.	8 hr.	10 hr.
Performance									
Tracking error	Static	Control	$\bar{X}$	5.43	5.21	4.74	4.51	4.77	4.73
			SD	1.71	.86	.52	.59	.41	.53
	Static	Alcohol	$\bar{X}$	5.29	5.92	5.86	5.78	4.74	4.84
			SD	1.46	1.15	1.01	1.28	1.74	.62
	Dynamic	Control	$\bar{X}$	6.73	5.44	5.06	4.88	4.91	4.50
			SD	1.86	.65	.65	.52	.51	.55
	Dynamic	Alcohol	$\bar{X}$	5.97	7.51	6.85	6.02	4.70	4.89
			SD	1.14	.82	1.13	.88	1.32	.54
Nystagmus									
Slow phase	Dynamic	Control	$\bar{X}$	5.42	3.30	3.29	3.06	2.35	3.10
			SD	2.83	2.16	1.81	2.24	1.37	1.54
Frequency	Dynamic	Control	$\bar{X}$	1.40	1.54	1.57	1.42	1.16	1.24
			SD	.91	1.05	.86	.94	.75	.73
Slow phase	Dynamic	Alcohol	$\bar{X}$	3.85	12.15	9.16	6.01	2.82	2.82
			SD	3.10	5.45	4.60	4.54	2.31	3.18
Frequency	Dynamic	Alcohol	$\bar{X}$	1.53	3.70	3.31	2.51	1.29	1.19
			SD	1.10	1.06	1.09	1.16	.83	.72

centages of increase or decrease in tracking error based on the predrinking level. Plots of the nystagmus measures across sessions are in Figure 2.

Control Group

The control group evidenced only a slight decline in static tracking error (an expected improvement with practice) from the pre through the 10-hr. session (see Fig. 1). None of the changes was statistically significant. However, dynamic tracking error evidenced a fairly steady decline from the pre through the 10-hr. testing sessions; all 1- to 10-hr. error scores were significantly ($p < .05$–$.01$) below the pre scores. Frequency measures of nystagmus for control *S*s showed no significant change across the six sessions, nor was there any significant difference in slow phase ocular velocity across the last five sessions. However, all but the last session of the latter were significantly below ($p < .05$–$.01$) the pretest level of slow phase activity (see Fig. 2).

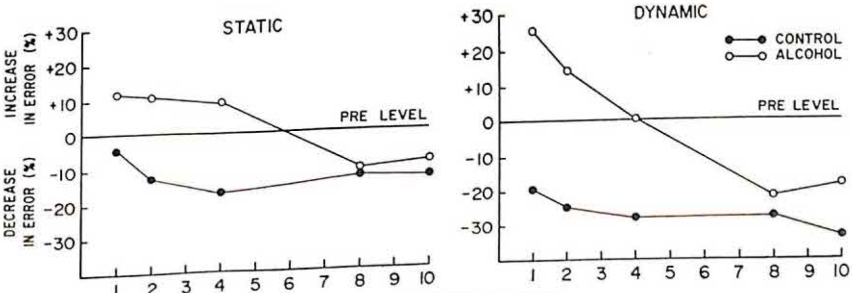


FIG. 1. Changes in tracking error under static (stationary) and dynamic (angular acceleration) conditions. (The O lines represent the base levels of tracking error during the preliminary trials. Tracking error scores for the five postdrinking sessions were converted to percentages of increase or decrease from the base levels.)

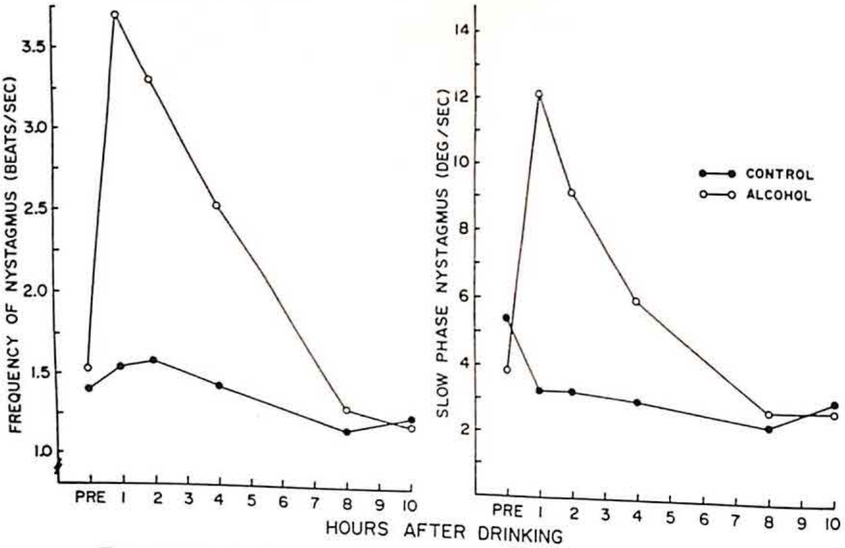


FIG. 2. Output of sampled nystagmus during each test session for the control and alcohol groups.

Alcohol Group

In contrast to control *S*s, both the average static and dynamic tracking errors increased for the alcohol group at the 1-, 2-, and 4-hr. testing sessions (see Fig. 1). However, only the increase for the 1-hr. static session ($p < .05$) and for the 1- and 2-hr. dynamic sessions ($p < .05$–$.01$) were significantly above the respective pre values. Measures of nystagmus for alcohol *S*s also presented a totally different picture from that of the control group (see Fig. 2). Both the degrees per second and the beats per second measures increased significantly ($p < .05$–$.001$) from the pretests through the 1-, 2-, and 4-hr. tests. At the 8- and 10-hr. tests, nystagmus was below the pre levels, but not significantly so.

Comparison of the Control and Alcohol Groups

In comparing the two groups, *t* tests were conducted on "change" (or difference) scores, that is, the difference in scores between the pre and the 1-hr. sessions, the pre and the 2-hr. sessions, etc.

Static tracking differences between the control and alcohol groups were significant ($p < .05$) only for the 4-hr. session. However, in the dynamic condition, differences between the two groups were significant ($p < .01$–$.001$) at the 1-, 2-, and 4-hr. sessions. Thus, with the addition of motion, the alcohol group performed with significantly more errors than the control group during the first 4 hr. following alcohol ingestion.

With respect to nystagmus, the control group had significantly less slow phase velocity ($p < .001$) and frequency of nystagmus ($p < .01$–$.001$) than did the alcohol group for the 1-, 2-, and 4-hr. sessions and less slow phase nystagmus ($p < .05$) for the 8-hr. session. Thus, the alcohol *S*s were less able than control *S*s to suppress their eye movements by fixation on the visual display during angular acceleration.

Discussion

Although alcohol effected an increase in tracking errors during static performance tests, the increase was significantly above predrinking levels only during the 1-hr. session. Moreover, differences between the alcohol and control groups in static tracking error were significant only during the 4-hr. session when the effects of alcohol were beginning to wane.

However, during vestibular stimulation, the eye-hand coordination required by the tracking task showed marked impairment by alcohol for the 1-, 2-, and 4-hr. sessions in comparison both with the alcohol group's predrinking performance and with the steady improvement demonstrated by the control group. The extent of this impairment appears to be directly correlated with the increased nystagmus activity to angular accelerations following alcohol ingestion. Thus, degradation of eye-hand coordination during stimulation of the semicircular canals appears to be closely related to the alcohol-induced loss of the ability to maintain adequate visual fixation (and visual acuity) on an object and thereby inhibits nystagmus. A similar degradation of visual acuity and tracking performance without alcohol has been reported previously (Gilson, Guedry, & Benson, 1970; Guedry, 1968). However, the magnitude of vestibular stimuli for commensurate losses was necessarily greater in those studies since, without alcohol, the visual fixation mechanism suppressing nystagmus was functioning normally.

These data have several practical implications. Activities that show little or no impairment in static situations following alcohol ingestion may be seriously degraded during motion. Further, the task required of the present *S*s was a relatively simple one, that is, *S* could concentrate on the single stimulus display. However, in many work activities, such as piloting an aircraft, the attention of the pilot has to shift from one stimulus "display" to another. It has been shown that deleterious effects of alcohol on performance in static situations are greatest when time-sharing of attention across several tasks is required (Chiles & Jennings, 1970) or if the task requires "divided attention" (Moskowitz & DePry, 1968). The addition of motion to a complex time-sharing task where performance is already degraded by alcohol might be especially hazardous. As such, in those situations which require the identification of unsafe levels of alcohol (e.g., in automobile driving) the basis for determining significant amounts of impairment should take into account the total array of stimuli that will impinge upon the individual.

As a final point, it should be noted that the average blood alcohol levels obtained in this study were considerably below most legal definitions of when an individual is presumed to be under the influence of intoxicating liquor (Erwin, 1970).[3] Only three *S*s exceeded a blood alcohol level of .090% during any of the blood-sampling periods.

---

[3] Motor vehicle laws in 23 states and the District of Columbia cite a blood alcohol level of .150% or more as a presumptive legal index; 21 states use .100%. Utah uses .080% as presumptive, and several other states are considering reducing their current levels to .080%. Five states have no defined levels.

REFERENCES

CHILES, W. D., & JENNINGS, A. E. Effects of alcohol on complex performance. *Human Factors*, 1970, **12**, 605–612.

COLLINS, W. E. Vestibular responses from figure skaters. *Aerospace Medicine*, 1966, **37**, 1098–1104.

ERWIN, R. E. Defense of drunk driving cases, criminal-civil. (2nd ed.) New York: Matthew Bender, 1970. (1970 cumulative supplement including appendix of pertinent statutes and supplemental index. Supplement prepared by L. A. Greenberg and Marilyn K. Minzer.)

GILSON, R. D., GUEDRY, F. E., & BENSON, A. J. Influence of vestibular stimulation and display luminance on the performance of a compensatory tracking task. *Aerospace Medicine*, 1970, **41**, 1231–1237.

GUEDRY, F. E. Relations between vestibular nystagmus and visual performance. *Aerospace Medicine*, 1968, **39**, 570–579.

HEIFER, U. Oculovestibuläre Reaktion und Blutalkoholkonzentration. *Hefte zur Unfallheilkunde*, 1969, **99**, 192–197.

MORGAN, C. T., COOK, J. S., CHAPANIS, A., & LUND, M. W. *Human engineering guide to equipment design.* New York: McGraw-Hill, 1963.

MOSKOWITZ, H., & DEPRY, D. Differential effects of alcohol on auditory vigilance and divided-attention tasks. *Quarterly Journal of Studies on Alcohol*, 1968, **29**, 54–63.

SCHROEDER, D. J. Influence of alcohol on vestibular responses to angular acceleration. *Aerospace Medicine*, 1971, **42**, 959–970. (a)

SCHROEDER, D. J. Alcohol and disorientation-related responses. I. Nystagmus and "vertigo" during caloric and optokinetic stimulation. (*FAA Office of Aviation Medicine Tech. Rep.* No. 71-6) Oklahoma City: Civil Aeromedical Institute, Federal Aviation Administration, 1971. (b)

(Early publication received May 6, 1971)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 564–569

INFLUENCE OF EXPECTATION OF TASK DURATION ON EFFICIENCY OF MUSCULAR ACTIVITY[1]

STJEPAN VIDAČEK[2] AND JULIUS WISHNER[3]

University of Pennsylvania

The purpose of this study was to test the hypothesis that intensity and organization of muscular activity is a function of *S*'s expectation of a task's duration. The electrical activity of the following muscles was measured while *S* was holding a 19-lb. weight under expectation for short and long tasks: biceps brachii, brachioradials, brachialis, and flexor carpi ulnaris. For the intervals measured, there was less electrical activity in all muscles during the long task. The less appropriate muscle showed the largest relative decrease. The results were interpreted as evidence of superior organization of activity under expectation of a task of long duration of work over expectation of a short duration of work.

It seems reasonable to suppose that foreknowledge of the duration of a task affects one's approach to it. For example, if *S* knows that he is to perform a task over a relatively long duration, he is likely to regulate his activity in such a way as to conserve some of his energy for later portions of the task. This can be manifested in a lower level of output per unit of time, or in a more economical or efficient expenditure of energy in the performance of the activity. In simple static work, for example, where *S* must keep a weight from dropping by flexing his forearm over a long period of time, expenditure of energy can be measured by means of electromyography, with electrodes attached to the skin above the relevant muscles.

There is a linear relation between the mechanical force of the contraction of a muscle and its electrical activity (Lippold, 1952). Thus, the electromyographic records of the electrical changes that accompany the contraction of the muscles can be used as an indicator of the level of muscle activity. If there is regulation of activity for the sake of energy conservation, it can be hypothesized that the level of the electrical activity of the muscles involved in static work will be lower for comparable periods of work if *S* knows or expects that he has to work for a long duration than if he expects to work for a short duration. Further, the decrease of electrical activity should be differential according to the degree of appropriateness of the various muscles to this type of task: namely, the decrease, with a set for a long duration, should be greater in less appropriate muscles than in more appropriate muscles (Wishner, 1955, 1962).

The task of static work selected for this experiment involved activity of the brachialis, the biceps brachii, and the brachioradialis, which are primarily concerned with flexion of the forearm (Goss, 1966; Romanes, 1966). Within the muscle system, the brachialis is the "workhorse" among the flexor muscles at the elbow. It is active especially during maintenance of specific flexed positions of the elbow, that is, isometric contraction in all positions of the forearm. This is not the case with the other two flexors of the forearm (Basmajian, 1967). With a large load, however, all three muscles are active during isometric flexion of the forearm (Bujas, Pavlina, Vidaček, & Vodanović, 1967).

The other muscles of the arm are also active during isometric flexion of the forearm under a large load, for example, the flexor carpi ulnaris (Bujas et al., 1967), which, although it is the primary flexor of the wrist, is only ancillary in flexion of the forearm (Romanes, 1966).

---

[1] The authors are grateful for the devoted assistance of Charles Maurer in running *S*s and analyzing data.

This study was carried out while the first author was a Public Health Service International Postdoctoral Research Fellow, under Grant No. TW01567 in the laboratories of the University of Pennsylvania. The research was supported for the most part by funds from Grant No. 04211 from the National Institute of Mental Health to the second author.

[2] Now head of the Laboratory for the Psychophysiology of Work at the Institute for Medical Research in Zagreb, Yugoslavia.

[3] Requests for reprints should be sent to Julius Wishner, Department of Psychology, University of Pennsylvania, 3813–3815 Walnut Street, Philadelphia, Pennsylvania 19104.

On the basis of these considerations, the specific hypotheses of this study are: (*a*) When *S* expects a task of long duration, the electrical activity of all the muscles is less than the electrical activity of these muscles during a comparable period of the same task when *S* expects to work for a short period; (*b*) The muscle action potentials of the flexor carpi ulnaris is reduced by a greater amount, relatively, in the long task than the electrical activity of the three other muscles.

METHOD

Subjects

The *S*s were 10 males, students and instructors at the University of Pennsylvania, ranging in age from approximately 20 to 45 yr.

Apparatus and Recordings

The overall structure of the apparatus in *S*'s room is shown in Figure 1. The *S* was seated in a modified dental chair. The tasks were performed on an apparatus consisting of a frame, which allows flexing of the forearm but prevents other arm movements. The frame was connected to a 19-lb. load through a pulley. The *S*'s task was to keep the weight above the floor and maintain his forearm at approximately a 90° angle at the elbow. At the front of *S*'s seat was a light bulb that went on when his forearm was at right angles to the upper arm. When the forearm was in any other position, the light went out, and *S* had to correct his position.

Recordings of integrated muscle action potentials during the tasks were made on a Grass Model 7 polygraph from surface electrodes, silver discs, 10 mm. in diameter, on the following four muscles of *S*'s right arm: biceps brachii, brachialis, brachioradialis, and flexor carpi ulnaris. The distance between each pair of electrodes was 1 in. The skin was appropriately prepared so that the resistance between pairs of electrodes was low (2,000–5,000 ohms) and similar. The raw signals were monitored on an osciloscope. The integrators were reset every 5 sec.

Procedure

There were four practice sessions on 4 days in the week preceding the experiment proper. The purpose of the practice sessions was to familiarize *S*s with the general conditions of the experiment and with the specific nature of the tasks. On each day of practice, each *S* had two sessions of static work: one for 30 sec., and one for as long as he could manage.

During the experiment proper, each *S* had a total of four static work sessions over 2 days. There were two sessions per day: one in the morning, the other in the afternoon. The order of the tasks was ABBA (A = 30-sec. task, B = work until subjective tolerance limit) for half the *S*s, and BAAB for the other half. The interval between the two sessions was 4 hr.

At the beginning of the fourth day of practice, *S*'s arm was prepared for recording of muscle action potentials during the task. The positions of electrodes were marked so they could be used during the experimental sessions the next day. At the beginning of the first session of the experiment proper, the electrodes were placed in the same position as during the last

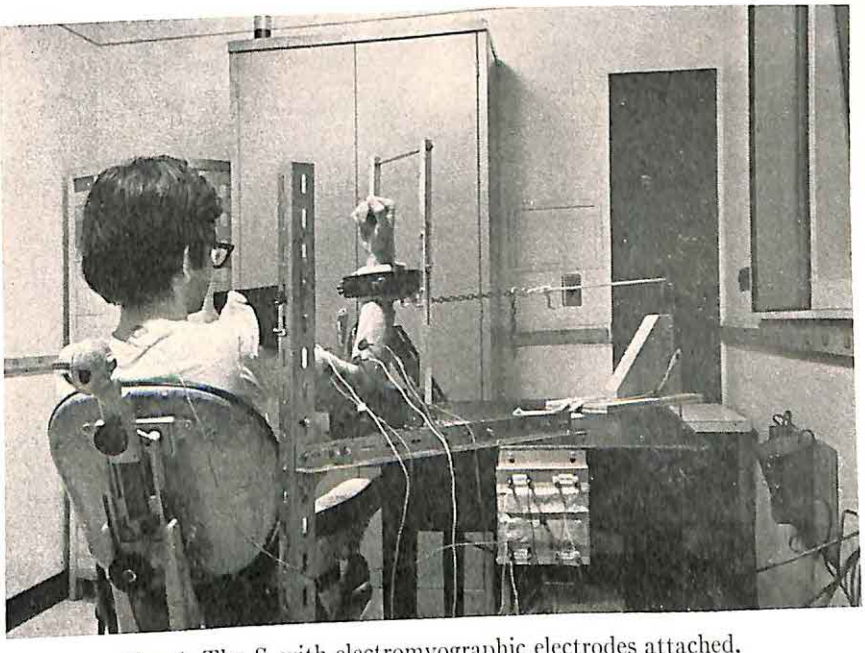


FIG. 1. The *S*, with electromyographic electrodes attached, holding the weight by means of a pulley.

practice session. A short period of calibration ensued, after which *S* began the first task. When the session was over, the electrodes were left on *S*'s arm, protected by an elastic bandage until the next session the same day. After the second session, the electrodes were removed, but their positions were marked.

The next day's procedure was the same as the first, except that the order of conditions was reversed.

Instructions

In the short duration task, *S* was told that he would have to keep the weight aloft for 30 sec. The apparatus was explained, and he was instructed to prepare for the task. Five sec. later, the task was begun. The *S* was told to keep the light bulb on by keeping his elbow at the appropriate angle. If the bulb went out, he was instructed to correct his position slowly. He was to maintain this position until *E* signalled the end of the 30-sec. period.

In the long task, *S* was told to hold the weight aloft as long as he could. He was warned that the task was difficult and resulted in somewhat painful muscle fatigue. He was implored to try his best, to hold the weight as long as he could in the prescribed position.

RESULTS

The *S*s had no difficulty in holding the weight in the prescribed position for the first 30 sec. The bulb did go out once in each of 6 sessions (of a total of 40 sessions): four times in the short task, for a total of 5 sec., and twice in the long task, for a total of 2.9 sec. The *S*s corrected their position smoothly, without significant effects on the recordings.

Figure 2 shows the mean integrated electrical activity of the muscles during the first 30-sec. period of static work. Except for the first 5-sec. period, the electrical activity of the muscles is similar from one 5-sec. period to the other. At the same time, the electrical activity of the muscles during the first 30 sec. of static work in the long task is lower in all muscles than the electrical activity of those muscles during the short task.

It can be seen that the electrical activity of the biceps brachii, and the brachioradialis registered by surface electrodes are higher than the electrical activity of the brachialis and the flexor carpi ulnaris. It is unreasonable, however, to regard this difference in the records as reflecting a real difference in the muscles' electrical activity. Various factors can attenuate the signals from the muscles to various extents, so that surface electrodes may give a very distorted picture of the underlying activity (Davis, 1959). For example, such factors as depth of the muscle, the position of electrodes in relation to the vector of electrical activity of the muscle, etc., can influence the voltages recorded, and the precise values of these influences are not easily measurable.

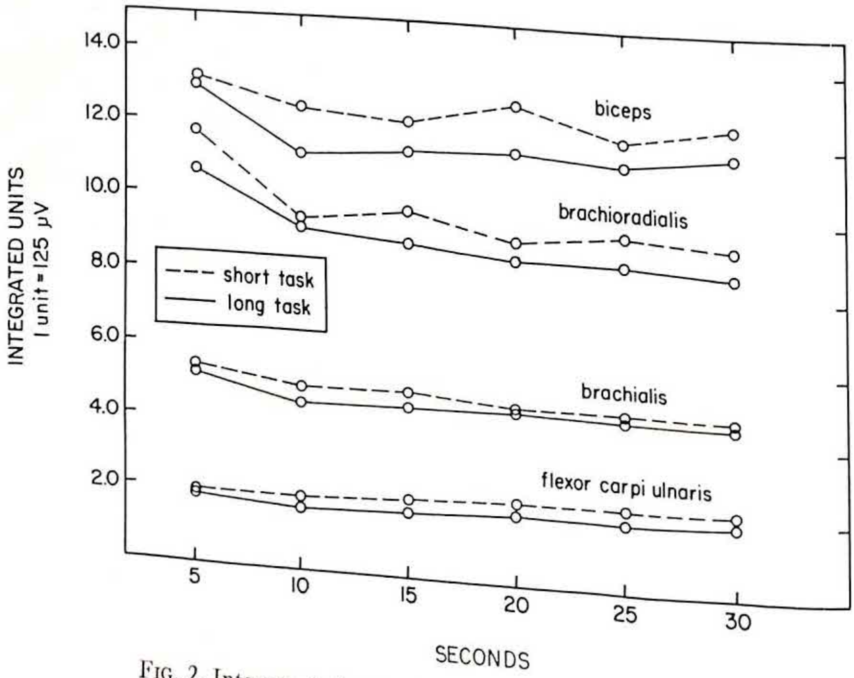


FIG. 2. Integrated electrical activity of the muscles as a function of duration of static work.

It is improbable that the main flexor of the forearm, the brachialis, is actually at as low a level of electrical activity in relation to the other two flexors as seems indicated in Figure 2. It seems much more likely that the recording of the electrical activity of the brachialis is attenuated to a greater degree by virtue of its depth and a number of other factors than the electrical activity of the other two muscles.

In this experiment we are more interested in the *changes* in electrical activity of the various muscles from one task to the other, and in the comparison of those changes, than in absolute values. The question is whether we can compare changes in electrical activity of various muscles, or, more specifically, whether the differences are distorted if they are derived from distorted values. Obviously, if the action potentials of one muscle are attenuated more than those of the other, then equal increases of electrical activity at the muscle will appear differentially smaller when they are picked up by surface electrodes. The attenuation of signals seems to be relatively large: less than .5 mm. distance between the fibre and its electrode reduces the muscle action potential by a factor of 10 (Buchthal, Guld, & Rosenfalck, 1954). Therefore, the absolute differences in signals picked up by surface electrodes give a distorted picture of changes in the muscle itself. Thus, the absolute differences cannot be used for the evaluation of changes in electrical activity of different muscles (Grossman & Weiner, 1966).

On the other hand, it is possible to assess the percentage of change of a muscle's electrical activity with a higher degree of validity. If the true electrical activity at the actual muscle site is decreased from one task to the other, it is probable that the percentage of the change produced at the recording site on the skin will be more comparable across muscles than changes in absolute values. Assuming that the major attenuating factors are multiplicative, as is indicated by previous research (Buchthal et al., 1954), this will hold even if the muscles being compared are at different depths.

In the following analysis, the integrated electrical activity of a muscle in the short task was taken as a base level and the integrated electrical activity of the same muscle during the first 30 sec. of work in the long task was taken as a percentage of the value of the electrical activity of the same *S* in the short task. These are the values shown in Table 1. They are the relative values of the electrical activity during the first 30 sec. of static work in the long task.

TABLE 1

VALUES OF THE INTEGRATED ELECTRICAL ACTIVITY OF MUSCLES DURING THE FIRST 30 SEC. OF THE LONG TASK, EXPRESSED AS A PERCENTAGE OF THE INTEGRATED ELECTRICAL ACTIVITY DURING THE SHORT TASK

S	Biceps	Brachialis	Brachio-radialis	Flexor carpi ulnaris	Σ	*M*
1	94.4	102.2	97.0	47.1	340.7	85.2
2	100.5	94.4	95.1	82.5	372.5	93.1
3	102.0	93.9	95.2	103.4	394.5	98.6
4	117.2	100.7	100.8	94.1	412.8	103.2
5	80.2	103.3	85.3	78.8	347.6	86.9
6	79.4	76.1	91.1	75.3	321.9	80.5
7	131.0	89.6	101.0	96.0	417.6	104.4
8	86.5	84.8	90.4	90.4	352.1	88.0
9	89.0	86.8	88.6	93.8	358.2	89.6
10	68.2	102.5	90.2	78.4	339.3	84.8
Σ	948.4	934.3	934.7	839.8	3657.2	914.3
M	94.8	93.4	93.5	84.0	365.7	91.4

Table 2 shows the mean relative values of the electrical activity during the first 30 sec. of the static work in the long task, *T* values of the Wilcoxon matched-pairs signed-ranks test, and associated *p* values.

The relative values of electrical activity of the main flexors of the forearm (the brachialis, the brachioradialis, and the biceps brachii) during the long task are similar, but only the decrease in the electrical activity of brachioradialis is statistically significant. With a one-

TABLE 2

MEANS OF RELATIVE VALUES OF INTEGRATED ELECTRICAL AACTIVITY OF MUSCLES DURING THE FIRST 30 SEC. OF THE LONG TASK, *T* VALUES OF THE WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST, AND ASSOCIATED *p* VALUES

Muscle	Relative value	*T*	*p*	
Biceps brachii	94.8	18		>.05
Brachialis	93.4	10	<.10	>.05
Brachioradialis	93.5	3		=.01
Flexor carpi ulnaris	84.0	1		<.01
M	91.4	5		=.02

tailed test, the relative value of electrical activity of the brachialis is also significant ($p < .05$).

With respect to the biceps brachii, it should be noted that variability is greatly elevated. It can be seen from the test-retest reliability that this variability does not reflect, in the main, underlying individual differences in biceps activity. Test-retest reliability of the electrical activity of biceps brachii is given by: $r_s = .50$ and $.55 (p > .05)$ for the long and short tasks, respectively.[4] The reliabilities of the other muscles are higher: for brachialis, $r_s = .90$ and .87; for brachioradialis, $r_s = .94$ and .78; and for flexor carpi ulnaris, $r_s = .87$ and .84 ($p < .01$ in all cases), for long and short tasks, respectively. The lability of the activity of the biceps is related to the variability in the activity of the other flexors during the same task, since their activity must be accommodated to the task of producing the same force during flexion of the forearm.

The largest relative decrease of electrical activity in the long task is shown in the flexor carpi ulnaris ($p < .01$). Although the absolute decrease of recorded electrical activity of the flexor carpi ulnaris is small (see Fig. 2), it should be noted that its activity in the short task, which is taken as the base level, is quite low to begin with. Floor effects operate against a substantial absolute decrease in the long task, and the decrease recorded thus takes on added significance, as represented in the relative, or percentage, decrease. The mean decrease in relative values of the electrical activity of all the muscles taken together is also statistically significant.

These results show that even if there are large individual differences in the interplay between the muscles, the general trend in the reduction of electrical activity during the first 30 sec. of work in the long task in relation to the electrical activity during a 30-sec. task can be described as follows: there is a reduction of electrical activity in all four muscles, but the most inappropriate muscle, flexor carpi ulnaris, shows the greatest reduction. At the same time, biceps barchii shows the greatest variability of changes in electrical activity from one condition to the other.

DISCUSSION

The results of this experiment show a significant reduction in the electrical activity of the muscles of an *S* who knows or expects that he has to work for a long period of time, as compared to the electrical activity of the muscles of an *S* who knows or expects that he has to work for a short period of time.

This decrease in the electrical activity of the muscles implies a decrease in the force of the muscles' contractions, if the assumption of a linear relationship between the force of contraction and the electrical activity that accompanies it is valid. However, it is difficult to give a quantitative statement of the amount of the reduction in the force of the muscles' contractions, especially since the slope of the regression of muscle action potentials upon force exerted is different for the same muscle among different *S*s (deVries, 1968). Similarly, the slope of the regression of muscle action potentials upon force of contraction is different for different muscles (Lenman, 1959). Nevertheless, the decrease of the electrical activity of the muscles in this experiment can be used as a general indicator of a true decrease in the activity of the muscles themselves, especially since it was possible to compare the electrical activity of the same muscle of each *S* in both conditions, that is, each *S* was his own control. Further, the electrodes were in identical positions in both conditions of the experiment during a single day, and in very similar positions from one day to another, and the apparatus was so constructed as to assure that the position of the limb and body were the same for all measurements.

It may seem paradoxical that there is differential force of contraction of muscles for identical loads. This becomes possible when it is realized that several muscles are active during an isometric contraction of the forearm. Some of them are more appropriate for this task, anatomically and physiologically, than others. The most appropriate muscle is brachialis. At the same time, the brachioradialis and the biceps help the brachialis to compensate for the external force of the load.

Some of the other muscles of the arm are active during static work. For example, flexor carpi ulnaris, which is the flexor of the wrist and only assists in flexing the forearm, shows considerable activity during static work, as does the main extensor of the forearm, the triceps. It is possible that at the beginning of a long static task, the activity of the muscles is more organized so that less appropriate and antagonistic muscles are not as active as during a short period of static work. In a long duration task, the main flexors of the forearm do not compensate for the force of the contraction of the antagonists as much as during a short task. At the same time, in the long task, reducing the activity of the less appropriate muscles, which produce little useful force relative to the level of their activity, makes the performance of the task more efficient.

---

[4] This finding may be interpreted as a warning against using muscle action potentials of the biceps as a general index of activity in the flexor muscle group of the forearm.

In sum, then, it appears that the static work task used in this experiment is a viable test of the hypothesis of conservation of effort and energy as a function of a set for differing durations of work. At the same time, it is essential to keep in mind that there was a set of background conditions that maintained *S*'s motivation for conservation. Thus, during the practice sessions, *S*s experienced the differences in difficulty and in effort required between the tasks. In the long duration task, there was discomfort and pain. Most *S*s were challenged and obviously put a great deal of effort into holding the weight as long as they could.

It is also possible that conservation of the activity of the muscles under the set for a long period of work is related to more intense motivation that could result in some dynamogenic changes in the vegetative nervous system. With higher motivation, the adjustment of the organism to the increased working requirements of the muscles may be improved, and the muscles may be able to perform the same work more economically (Bujas & Petz, 1952).

A related study is that of Walster and Aronson (1967) on the effect of expectancy of a task's duration on the experience of fatigue. They found that those *S*s who expected to continue a series of fatiguing tasks reported less fatigue than *S*s who were led to believe that their chore was almost at an end. The authors believe that *S*s suppress feelings of fatigue until their task is virtually completed. Our results show that expecting a long task gives rise to economy of effort and energy from the very beginning of the task. Thus, there would be not only suppression of the feeling of fatigue, as shown in verbal reports, but also, if an objective index can be found, an actually lower level of fatigue for any given time, as a function of expectancy for a long duration of work. Indeed, it should be possible in future experiments to explore the possibility that the developing patterns of muscle activity are related to reports of fatigue.

REFERENCES

BASMAJIAN, V. J. *Muscles alive.* (2nd ed.) Baltimore: Williams & Wilkins, 1967.

BUCHTHAL, F., GULD, C., & ROSENFALCK, P. The origin of the motor unit potential. *Acta Physiologica Scandinavica*, 1954, **31**, 7–8 (Suppl. 114).

BUJAS, Z., PAVLINA, Ž., VIDAČEK, S., & VODANOVIĆ, M. Electromyographic investigation of fatigue and recovery in the field of static work. *Acta Instituti Psychologici Universitatis Zagrabiensis*, 1967, **51**.

BUJAS, Z., & PETZ, B. L'influence des stimulants psychiques sur le rendement du travail et sur la consommation d'oxygène au cours du travail statique. *Arhiv za higijenu rada*, 1952, **3**, 282–295.

DAVIS, J. F. *Manual of surface electromyography.* (WADC Tech. Rep., 59–184) 1959. Wright-Patterson Air Force Base: Ohio (Project No. 7184, Task No. 71580).

DEVRIES, A. H. "Efficiency of electrical activity" as a physiological measure of the functional state of muscle tissue. *American Journal of Physical Medicine*, 1968, **447**, 10–22.

GOSS, M. C. (Ed.) *Gray's anatomy of the human body.* Philadelphia: Lea & Febiger, 1966.

GROSSMAN, I. W., & WEINER, H. Some factors affecting the reliability of surface electromyography. *Psychosomatic Medicine*, 1966, **28**, 78–83.

LENMAN, R. A. J. Quantitative electromyographic changes associated with muscular weakness. *Journal of Neurology, Neurosurgery, and Psychiatry*, 1959, **22**, 306–310.

LIPPOLD, O. C. J. The relation between integrated action potentials in a human muscle and its isometric tension. *Journal of Physiology*, 1952, **117**, 492–499.

ROMANES, J. G. *Cunningham's manual of practical anatomy.* (13th ed.) Vol. 1. *Upper and lower limbs.* London: Oxford University Press, 1966.

WALSTER, B., & ARONSON, E. Effect of expectancy of task duration on the experience of fatigue. *Journal of Experimental Social Psychology*, 1967, **3**, 41–46.

WISHNER, J. The concept of efficiency in psychological health and in psychopathology. *Psychological Review*, 1955, **62**, 69–80.

WISHNER, J. Efficiency: Concept and measurement. In *Personality research*, Proceedings of the Fifteenth International Congress of Applied Psychology, Vol. 2., Copenhagen: Munksgaard, 1962.

(Received September 15, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 570–579

PRICE, BRAND NAME, AND PRODUCT COMPOSITION CHARACTERISTICS AS DETERMINANTS OF PERCEIVED QUALITY[1]

JACOB JACOBY,[2] JERRY C. OLSON, AND RAFAEL A. HADDOCK

Purdue University

While price and brand image have both been found to be determiners of product quality perception, the potency of these two cues has never been directly compared. Moreover, those studies which found price to be a determiner of perceived quality manipulated only price information, without permitting actual composition characteristics to vary across brands. A 2 × 2 × 2 × 3 factorial experiment, using 136 adult male beer drinkers, and four test beers, examined the effects of price, composition differences, and brand image cues on the perception of beer quality. Price was found to serve as an indicant of product quality when it was the only cue available but not when embedded in a multicue setting. Brand image had a stronger effect upon quality perception, particularly for brands with strong positive images. In addition, it was found that neither price nor brand name had significant effects on perceived quality except when product composition characteristics were allowed to vary between product samples. Last, in contrast to earlier findings, the data suggest that beer drinkers possess at least some ability to distinguish among different brands of beer on the basis of composition (i.e., taste and aroma) cues alone.

As the literature on perceived risk (cf. Cox, 1967) clearly demonstrates, consumer purchase decisions are frequently made under conditions of varying uncertainty regarding the product and its attributes. To reduce such uncertainty, consumers seek and process information regarding the product and generally attempt to form accurate impressions of it. Given that products may be viewed "as an array of cues," the "consumer's task in evaluating a product is to use cues from the array as the basis for making judgments about the product (Cox, 1962, p. 413)."

One impression usually of considerable importance to the consumer is the product's (or brand's) quality. Cues relevant to forming impressions of quality include (*a*) price; (*b*) product composition characteristics such as taste, aroma, color, style, and size; (*c*) packaging; (*d*) brand, manufacturer (i.e., corporate), and store image; (*e*) advertising; (*f*) word-of-mouth reports; and (*g*) past purchase experience. Of these cues, greatest attention thus far has been directed toward price. Shapiro (1968) noted that "price is a powerful piece of information for the consumer. . . . Since price is concrete and measurable, the consumer views it with much confidence. He trusts it more than most cues concerned with quality (p. 20, 22)."

The first conceptual treatment of the price-quality relationship is usually credited to Scitovsky (1967), and subsequent empirical investigations (Gabor & Granger, 1966; Leavitt, 1954; McConnell, 1968a, 1968b; Tull, Boring, & Gonsior, 1964) have found that, with other cues held constant, price is indeed a determiner of perceived quality. Illustrative of this approach is McConnell's work. Labeling different bottles of the same brand of beer M, L, P, and associating different prices with these labels ($1.30, $1.20, and 99¢ per six-pack, respectively), he was able to demonstrate that the high-priced brand was perceived to be of higher quality than the medium-priced brand ($p = .06$). There was no difference in perceived quality, however, between the medium- and low-priced brands.

As Enis and Stafford (1969) noted, ". . . it seems fairly clear that if the only information consumers are given about product quality is price information, they will indeed judge quality by price (p. 340)." Combining price and store image cues in one experiment, these investigators still found price to strongly affect quality perception (Stafford & Enis, 1969).

---

[1] The authors gratefully acknowledge the critical reading and helpful comments provided by Howard Fromkin.

The authors wish to thank William H. Campbell, Kevin C. Peterson, and Raymond C. Stokes for their assistance in conducting the experiment.

[2] Requests for reprints should be sent to Jacob Jacoby, Department of Psychology, Purdue University, Lafayette, Indiana 47907.

It is noteworthy that nowhere, in all those studies that placed *S*s in choice situations involving real products, have actual differences in product composition been allowed to vary. Investigators have given *S*s identical samples of beer, or carpet swatches cut from the same roll, and simply attributed different labels, prices, and images to them. In contrast, and in consonance with real-world situations where product differences can and do vary, the present investigation allowed actual physical differences to vary, while also manipulating price and brand image cues. Both product composition differences and price may be considered objective attributes of the product, while brand image may be defined as the subjective, emotional cluster of meaning and symbols that the consumer attributes to particular brands. Usually brand names are sufficient to evoke brand images.

Several considerations suggested beer as the product to be studied. First, a replication of McConnell's (1968a, 1968b) often cited findings, in which price was the only cue to vary, was considered desirable. A second purpose was to examine whether beer drinkers could discriminate among different brands on the basis of taste and aroma cues alone, given a range encompassing substantial differences in product composition. Earlier work had indicated that consumers are unable to distinguish among different brands of beer in "blind" taste tests (Allison & Uhl, 1964). Reexamination of similar data from cigarettes and cola studies, however, yielded evidence to suggest some ability to discriminate among various brands in these product classes (Myers & Reynolds, 1967, p. 18). Therefore, examination for similar abilities with respect to beer seemed worthwhile.

With price absent, and given limited (or in-) ability to discriminate among different beers on the basis of taste and aroma cues, what does determine perception of product quality? Several investigators (Cox, 1962, Froman, 1953; White, 1966) have suggested that nonobjective criteria, especially various types of images (e.g., brand, store, corporate, product, and salesman), are frequently used by consumers to evaluate alternative brands. Empirical support was provided by a second part of the Allison and Uhl (1964) study in which *S*s received properly labeled bottles of beer and were again asked to make the same judgments regarding several product attributes including overall quality. The results of this manipulation clearly indicated that

> labels, and their associations, did influence their evaluations. In other words, product distinctions or differences, in the minds of the participants, arose primarily through their receptiveness to the various firms' marketing efforts rather than through perceived physical product differences (p. 39).

Given that price and brand image have each been demonstrated to affect perception of beer quality, the present experiment compared the relative efficacy of these two cues.

In summary, the general purpose of this study was to examine the unique and interaction effects of three cues—price, composition differences, and brand image—on perception of brand quality. Specific foci of interest were: (*a*) to determine whether the price-quality relationship could be replicated in a more complex, multicue environment; (*b*) to examine the relative efficacy of price and brand image as determinants of perceived quality; and (*c*) to determine whether beer drinkers are able to discriminate among different brands of beer on the basis of only taste and aroma cues, given substantially different beer samples.

METHOD

Subjects

The *S*s, 136 male volunteer students who answered an advertisement in the campus newspaper requesting participants for a "beer taste-testing study," were all over 21 yr. old, and each stated that he was a frequent beer drinker and had consumed at least 12 oz. of beer during the preceding week. Potential *S*s called a telephone answering service in response to the ad, and those that met the age and usage criteria were randomly assigned to one of the eight experimental groups.

Design

A 2 × 2 × 2 × 3 factorial experiment with 17 *S*s per cell was conducted which consisted of two levels of price (present, absent), brand name (present, absent), and composition differences across beer samples (present, absent) as between-*S* factors, and three beer samples as within-*S* factors. The dependent variable of interest was perceived overall quality, and repeated measures were taken over three beer samples for each *S*.

Instruments

Each *S* received a questionnaire booklet that gathered elevant information concerning his beer-drinking habits and familiarity with various brands of beer, explained and administered the appropriate manipulations, and collected five dependent measures (perceptions of aroma, bitterness, strength, smoothness, and overall quality) for each of the three beer samples. The first four dependent measures were included to disguise the purpose of the experiment.

There was a graphic 0–100-point scale containing 10 equally spaced, numbered graduations and having endpoints labeled "worst beer" and "best beer" for each dependent variable. Written and verbal instructions directed *S*s to evaluate the beer samples served to them in relation to the best and worst beers they had ever tasted. Evaluations were recorded by placing three arrows at appropriate points on each scale and labeling each with the name or letter of the beer sample being rated.

Procedure

Each of the eight cells was run separately. The *S*s were seated at a long table and separated by table-top partitions to cut off between-*S* communications. A "credible-appearing" source in his early 40s, who served as *E*, introduced the experiment as a marketing study funded by a national brewer for the purpose of developing reliable instruments to measure beer taste perceptions. The *S*s were told that college students were large consumers of beer, and therefore the manufacturer was very much interested in their responses.

The *S*s were then given the questionnaire booklet and completed the first two pages regarding their beer-drinking habits and general familiarity with beer brands. Upon completion, *E* set up the appropriate manipulations and, using a blackboard, demonstrated how to use the graphic rating scales.

Four different brands of beer were used. A nationally distributed ultrapremium beer (costing $1.60 per warm six-pack), a nationally distributed popular-priced beer (costing $1.35), and a regionally distributed inexpensive beer (costing 99¢) were used for the 68 *S*s in the four "composition differences present" cells. These beers were labeled Y, X, and Z, respectively, in the "brand name absent" cells. The three brands were selected from among 20 different brands on the basis of a pretest that indicated that a comparable group of *S*s could, relative to other beers, accurately estimate the actual cost per warm six-pack of these brands, with minimal variance in these estimates across *S*s. These selection criteria had two purposes: (*a*) to minimize differential cross-*S* familiarity with respect to the beers, a variable likely to influence preference ratings (cf. Zajonc, 1968); and (*b*) to cover as broad a range of domestic beers (in terms of price, brand image, and probable composition differences) as possible. Another nationally distributed popular-priced beer, whose price was also easily and accurately estimated, was used for all three beer samples presented to the 68 *S*s in the four "composition differences absent" cells (see Table 1).

For those cells in which neither price nor brand names were provided, *S*s were simply told they were evaluating three beers generally available in the local area and here designated X, Y, and Z. When price was given, the beers were listed in the questionnaire booklet with their actual store prices (per six-pack) appearing next to them. The same procedures were used when brand name (used to evoke brand image) was present except that, instead of X, Y, and Z, the actual brand names of the beer samples were provided.

For the four cells in which the same brand of beer was used for all three beer samples, directions were identical to those given when the beer samples differed. That is, *S*s in those cells were also led to believe that they were tasting three brands of beer readily available in the local area.

Beer samples were given to each *S* in plastic beige-colored cups that tended to disguise color cues. As appropriate, the cups were marked either with brand names or letters, and the accompanying instructions (in the experimental booklet) did or did not present information. The *S*s were told to sample the beers in any order but to complete each dependent measure rating scale for all three beers before proceeding to the next scale, for example, complete the aroma ratings for all three beers, then proceed to the bitterness scale, etc. Because of university regulations, 2 oz. of beer were served in each cup, making a total of 6 oz. per *S*.

Precautions were taken to insure that the beer samples were quite cold and had little foam. The three samples for each *S* were poured at the same time (out of sight of *S*s) and immediately taken to him. Comments elicited at the conclusion of the experiment indicated that the state each sample arrived in was comparable to normal drinking conditions.

RESULTS

Two general assumptions should be met for the repeated-measures analysis of variance model to hold: (*a*) homogeneity of the variance-covariance matrixes and (*b*) symmetry of the pooled variance-covariance matrix (see Winer, 1962, pp. 369–374). Tests conducted for both these assumptions were not statistically significant ($\chi^2 = 51.80$, $df = 42$, and $\chi^2 = 5.69$, $df = 7$, respectively, both $ps > .10$), and therefore the model was assumed appropriate for these data.

Table 1 presents the means and standard deviations of perceived quality ratings for each experimental cell. Simple inspection of these data reveals at least two items of interest. First, there was substantially greater variability of interbeer sample quality ratings when composition differences were present than when the samples were of the same beer. In fact, the range of the overall beer sample means was twice as great when composition differences were present (70.88–37.58 = 33.30

TABLE 1

MEANS AND STANDARD DEVIATIONS OF OVERALL QUALITY RATINGS

Condition	Composition differences present (samples of different beers)			Composition differences absent (samples of same popular-priced beer)		
	Beer Z (regional inexpensive)	Beer X (popular priced)	Beer Y (ultra popular priced premium)	Beer Z (purported regional inexpensive)	Beer X (purported popular priced)	Beer Y (purported ultra-premium)
Brand name present						
Price present						
$\bar{X}$	33.47	65.71	79.82	50.76	64.18	76.18
SD	17.82	23.65	13.21	19.49	16.47	17.96
Price absent						
$\bar{X}$	34.18	64.88	71.41	50.12	56.88	71.06
SD	17.23	15.81	15.37	19.99	23.33	13.65
Brand name absent						
Price present						
$\bar{X}$	34.47	66.12	67.65	55.35	54.65	68.41
SD	18.08	17.90	22.84	21.08	21.79	21.25
Price absent						
$\bar{X}$	48.18	55.12	64.65	51.47	68.53	57.18
SD	22.54	22.99	16.10	21.66	20.38	19.30
Overall						
$\bar{X}$	37.58	62.96	70.88	51.93	61.06	68.21
SD	19.60	20.42	17.82	20.21	20.95	19.17

Note.—n = 17 per cell.

vs. 68.21–51.93 = 16.28), with the regional inexpensive beer's low ratings accounting for most of this difference. Second, the brand name (BrN) absent, price (P) present, compositional differences (CD) absent cell replicated McConnell's (1968a, 1968b) findings, that is, the high-priced brand was perceived to be of higher quality than the medium- and low-priced brands ($t = 1.85$, $df = 16$, $p < .05$ and $t = 1.81$, $df = 16$, $p < .05$, respectively), with no significant difference between the latter two.

Results of a four-way ANOVA based on the data presented in Table 1 are summarized in Table 2. A strong beer samples (BS) main effect was obtained ($F = 57.99$, $df = 2, 256$, $p < .0001$), and a Newman-Keuls analysis indicated that all differences between pairs of means for the three beers ($\bar{X}_z = 44.75$, $\bar{X}_x = 62.01$, $\bar{X}_y = 69.54$) were significant beyond the .01 level. These differences were all in the expected direction, with the ultrapremium beer (Y) receiving the highest mean quality rating and the regional inexpensive beer (Z) receiving the lowest mean rating.

Two two-way interactions were significant in the overall ANOVA, CD × BS and BrN × BS, and are shown in Figure 1 (Parts a and b, respectively).

TABLE 2

ANALYSIS OF VARIANCE OF OVERALL QUALITY RATINGS FOR COMPLETE DESIGN

Source	*df*	*MS*	*F*
Between *S*s			
Price (P)	1	378.55	1.02
Brand name (BrN)	1	511.88	1.38
Composition differences (CD)	1	1083.88	2.92
P × BrN	1	285.00	<1
P × CD	1	21.20	<1
BrN × CD	1	.06	<1
P × BrN × CD	1	9.12	<1
Error (between)	128	371.06	<1
Within *S*s			
Beer samples (BS)	2	21972.33	57.99***
P × BS	2	762.67	2.01
BrN × BS	2	2015.99	5.32**
CD × BS	2	3143.13	8.30**
P × BrN × BS	2	87.90	<1
P × CD × BS	2	756.71	2.00
BrN × CD × BS	2	234.74	<1
P × BrN × CD × BS	2	1461.18	3.86*
Error (within)	256	378.88	

* $p < .025$.
** $p < .01$.
*** $p < .0001$.

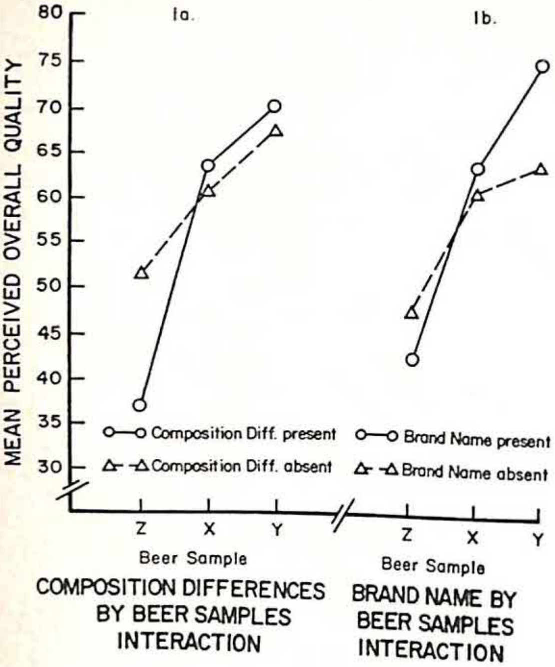


FIG. 1. Two interactions with beer samples from overall analysis of variance.

Analysis of variance of the simple main effects indicate that the effect of composition differences (see Part a, Figure 1) was significant only for Beer Z ($F = 18.62$, $df = 1, 384$, $p < .01$), and the effect of brand name (see Part b, Figure 1) was significant only for Beer Y ($F = 9.30$, $df = 1, 384$, $p < .01$). It appears that, for the three brands of beer used, presence of the true compositional characteristics does not significantly affect one's perception of their quality, except for the low-priced regional beer (Z), in which case presentation of the actual beer depressed the perceived quality ratings. One might say that Beer Z tastes worse than its image (or at least worse than the fourth beer used in CD absent conditions). Similarly, it appears that only the high-priced ultrapremium beer's (Y) image (as mediated by brand name) was strong enough to significantly increase perceptions of quality. As might be expected, Beer Z seems to have a somewhat negative image, since presentation of its brand name depressed perceived quality ratings, although not to a significant extent ($F = 2.48$, $df = 1, 384$, $p > .20$). Brand name had no significant effect on Beer X.

The four-way interaction, P × BrN × CD × BS, was also significant ($p < .025$) and is shown in Figure 2. The increase in variability under CD present conditions is clearly obvious. To clarify interpretation, simpler levels of analysis were sought by analyzing four three-way ANOVAs.

By dividing the study into two three-way analyses—one in which CD are absent and one in which they are present—one obtains a design similar to most of the previously conducted quality perception research in the former and an extension of such studies by allowing interbrand composition differences to vary as they do in a natural setting in the latter. Results of these two analyses are presented in Table 3.

The only significant effect for CD absent was the BS main effect. All intermean differences ($\bar{X}_z = 51.93$, $\bar{X}_x = 61.06$, $\bar{X}_y = 68.21$) were significant beyond the .05 level using a Newman-Keuls analysis. A logical and parsimonious explanation for these differences is that they are due to demand characteristics (cf. Orne, 1962) created by the experimental procedures. Given that these *S*s (under CD absent) were told that they were evaluating three different beers and that *S*s in three out of these four cells received additional information (i.e., brand name and/or price) sup-

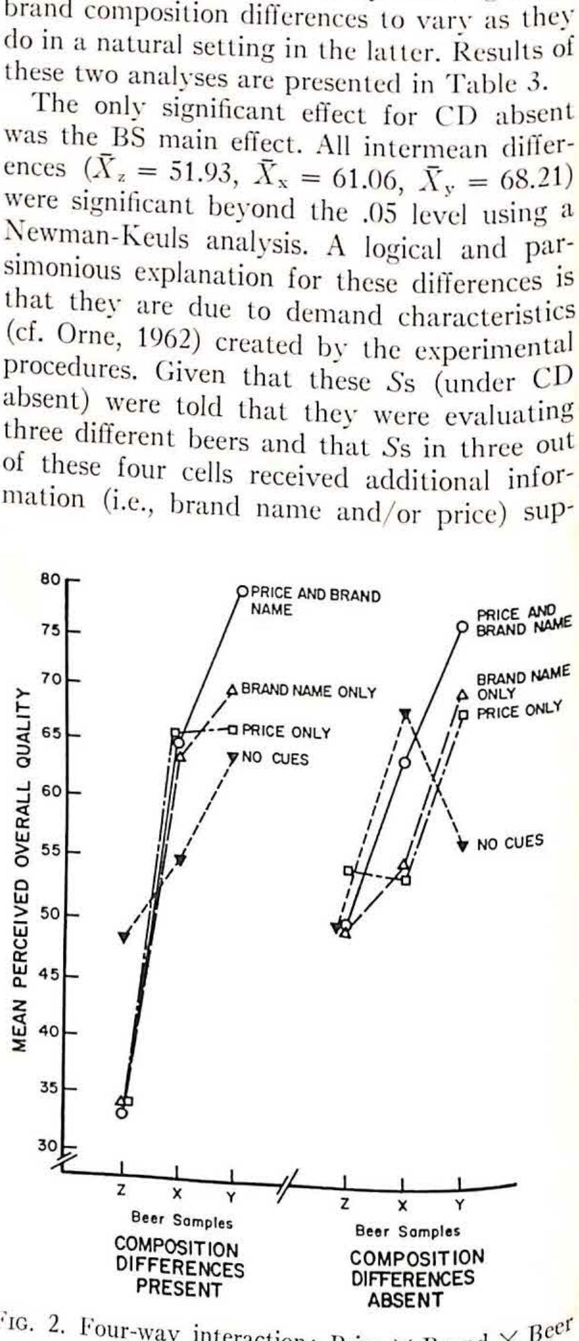


FIG. 2. Four-way interaction: Price × Brand × Beer Sample × Composition Differences.

TABLE 3

ANALYSIS OF VARIANCE OF OVERALL QUALITY RATINGS FOR TWO LEVELS OF COMPOSITION DIFFERENCES

Source	*df*	Composition differences absent		Composition differences present	
		MS	*F*	*MS*	*F*
Between *S*s					
Price (P)	1	289.46	<1	110.29	<1
Brand name (BrN)	1	261.57	<1	250.37	<1
P × BrN	1	189.04	<1	96.08	<1
Error (between)	64	369.48		372.64	
Within *S*s					
Beer samples (BS)	2	4527.66	11.13***	20587.80	58.64***
P × BS	2	559.37	1.38	960.01	2.73*
BrN × BS	2	946.51	2.33	1301.21	3.71**
P × BrN × BS	2	955.69	2.35	593.39	1.69
Error (within)	128	406.65		351.11	

* $p \cong .07$.
** $p < .03$.
*** $p < .0001$.

porting these cues, the average *S* tended to rate the quality of his three beer samples in a way consistent with these demand characteristics. In other words, he rated the supposed ultrapremium, high-priced beer (Y) relatively high, the popular-priced beer (X) lower, and the regional inexpensive beer (Z) lowest. In contrast to earlier studies, however, price exerted no effect upon quality perception.

For the ANOVA for CD present (see Table 3), the main effect due to BS was again significant ($F = 58.64$, $df = 2, 128$, $p < .0001$), and a Newman-Keuls analysis revealed that differences in mean quality ratings between Beers Z and X, and Beers Z and Y were significant beyond the .01 level, with the difference between Beers X and Y significant beyond the .05 level ($\bar{X}_z = 37.57$, $\bar{X}_x = 62.96$, $\bar{X}_y = 70.88$). This main effect was much stronger in this instance than when CD was absent, indicating the amount of increased variability due to the presence of actual composition differences (e.g., taste, carbonation, bitterness, etc.) between the beer samples above that due to the demand characteristics.

The BrN × BS interaction was also significant ($p < .03$) under CD-present conditions. Again, only the simple, simple main effect of brand name for the high-priced, ultrapremium beer (Y) was significant ($F = 4.26$, $df = 1, 192$, $p < .05$), indicating that brand image positively affected quality perceptions of this beer but had no significant effect on the other two beers.

Although the P × BS interaction (at CD present) approached but did not achieve a customary level of statistical significance ($F = 2.73$, $df = 2, 128$, $p \cong .07$), it was examined further since it was of a priori importance. Subsequent ANOVA of the simple, simple effects of price at the three levels of BS failed to reveal any statistically significant effects (all ps $> .20$). In other words, knowledge of price (at this level of analysis) had no significant effect on perceived quality ratings for any of the three real beers.

A last set of three-way analysis that, with its relevance to prior studies, seems to be most generalizable is for both levels of BrN (i.e., BrN present and BrN absent). Results of these two ANOVAs are presented in Table 4. A significant main effect of BS occurs under both levels of BrN (both ps $< .0001$). As might be expected, this main effect is stronger when the beer samples are identified by their brand names ($F = 60.83$ vs. $F = 12.26$, $df = 2, 128$), as evidenced by the intermean differences for BrN present ($\bar{X}_z = 42.13$, $\bar{X}_x = 62.91$, $\bar{X}_y = 74.62$) compared to those for BrN absent ($\bar{X}_z = 47.31$, $\bar{X}_x = 61.10$, $\bar{X}_y = 64.47$). All intermean differences within both sets are significant beyond the .01 level

TABLE 4

ANALYSIS OF VARIANCE OF OVERALL QUALITY RATINGS FOR TWO LEVELS OF BRAND NAME

Source	*df*	Brand name absent		Brand name present	
		MS	*F*	*MS*	*F*
Between *S*s					
Price (P)	1	3.31	<1	660.42	1.74
Composition differences (CD)	1	533.82	1.47	550.12	1.45
P $\times$ CD	1	1.26	<1	29.06	<1
Error (between)	64	362.70		379.41	
Within *S*s					
Beer samples (BS)	2	5581.78	12.26***	18406.53	60.83***
P $\times$ BS	2	651.68	1.43	198.89	<1
CD $\times$ BS	2	1078.72	2.37	2299.15	7.60**
P $\times$ CD $\times$ BS	2	2116.50	4.56*	101.39	<1
Error	128	445.16		302.60	

* $p < .02$.
** $p < .002$.
*** $p < .0001$.

using a Newman-Keuls analysis, except the nonsignificant X–Y difference for BrN absent.

A different perspective on the interesting phenomonon is reflected in the two analyses shown in Table 4. CD do not significantly (i.e., with a $p < .05$) interact with BS unless another cue (e.g., BrN or P) is present. The significant P $\times$ CD $\times$ BS interaction ($p < .02$) at BrN absent and the CD $\times$ BS interaction ($p < .002$) at BrN present both provide evidence to support this statement. To be more specific, the CD $\times$ BS interaction is not significant when BrN is absent ($p \cong .095$), but under BrN present conditions, the CD $\times$ BS interaction is significant beyond the .002 level. Probing of this latter interaction reveals that the effect of CD is again significant only for Beer Z ($F = 14.73$, $df = 1, 98$, $p < .01$).

Likewise, probing of the three-way interaction, P $\times$ CD $\times$ BS, reveals a significant CD $\times$ BS interaction only when price cues are present ($F = 5.04$, $df = 2, 64$, $p < .009$). The CD $\times$ BS interaction when no cues were present (i.e., neither P nor BrN) was not significant ($F = 2.01$, $df = 2, 64$, $p \cong .14$). Therefore, the results of this experiment indicate that composition differences have a significant effect on perceived quality ratings only when supported by another cue or, stated conversely, cues such as brand name have an effect only when product composition differences are allowed to vary (e.g., see interaction under CD present, Table 3).

A summary of the basic findings includes the following: Composition differences did exert a significant effect upon quality perception, but only for Beer Z, in which case using the "real" beer depressed quality ratings below those obtained when using the "fake" beer. Similarly, brand image exerted a significant effect upon quality perception, but only for Beer Y, the most expensive and status-oriented beer. Price, except in isolation, did not exert any significant effect upon quality perception. In other words, for the three brands of beer and the sample of *S*s tested, brand name and actual composition characteristics were more important determiners of perceived quality than was price. Last, there was substantially greater variability in quality ratings when composition differences were allowed to vary, thereby suggesting that beer drinkers can make some distinctions among different brands on the basis of composition (i.e., taste and aroma) cues.

DISCUSSION

In light of previous research on the price-quality relationship, one of the most striking aspects of the obtained results was the inability of the price cue (when combined with other cues) to significantly influence perception of product quality. While McConnell's (1968a, 1968b) findings were replicated, it would appear that when price is presented in conjunction with other cues potentially indicative of product quality, it is used to a much lesser extent than when presented alone. It should be noted, however, that price was not manipulated as in earlier studies (i.e., by assigning different price levels to the same beer). Inasmuch as the purpose of the present research was primarily to determine the effect of price relative to other cues in a multicue setting, not to examine whether different prices would affect quality perception, the appropriate manipulation was simply the presence or absence of price.

Comparing the relative strength of price and brand-image cues, the results indicate that, while the effect of price was not significant, awareness of the brand name did influence quality perception. A priori, one would expect knowledge of the brand name to increase perception of quality for the ultrapremium beer and decrease it for the inexpensive regional beer, and this was precisely the trend observed (see Table 1).

It would therefore seem as if brand image is a more important cue than price in determining quality perception of beer. The price-brand name distinction is somewhat confounded. However, especially for established products and knowledgeable *S*s, knowing only that the brand name provides more information than simply knowing the price. Again, this is reflected by the data (see Table 1). The presence of brand name alone produced results more similar to the brand name-price present condition than did the presence of price alone. This confound was virtually insured by the pretest criteria used to select the experimental brands.

The variables that affect quality perception seem to manifest themselves primarily through interaction with other variables. Of the four variables considered, only one (beer samples) was strong enough to consistently affect quality perception across a variety of situations. The effects of the other variables were moderated by the situation in which they were introduced. This tendency toward interactive effects may limit the number of variables that can be studied in a single experiment, since higher order interactions become exceedingly difficult to interpret.

Considering the nature of these interactions more closely, one of the most interesting, and potentially important, results of the present study is the lack of Cue × Product Sample interactions when differences in brand composition were *not* present, that is, when the samples were identical. Significant Cue × Product Sample interactions occurred only when actual composition differences existed between the three samples, that is, when they were really different beers.

If this rather strong result is reliable, it has important implications for future research. Since past experimental investigations of quality perception have manipulated other cues while holding actual differences between the samples constant (i.e., nonexistent), they may be using an oversimplified representation of reality and, therefore, missing important and complex relationships between the cues and the brands themselves. Consequently, relationships obtained under these conditions may have little or no generalizability to real world considerations.

One of the more interesting findings was the evidence to indicate that beer drinkers are capable of discriminating quality differences in the expected direction, given only taste and aroma cues. If one assumes that, at least for the brands studied, differences in price and image reflected actual differences in quality, then perceived quality corresponded to objective quality for these *S*s.

Although it did tend to confound some of the results, another interesting finding was a "letter effect" which evidenced itself in the price absent- brand name absent- composition differences absent cell. In some respects, this corresponds to Tucker's (1964) investigation in which housewives selected one out of four loaves of bread, differentiated only by the letters H, L, M, and P, over 12 trials. Despite the fact that all four loaves were identical, were taken from a single oven on the day of delivery, and, except for the letter labels, were packaged identically, half the housewives in the study manifested brand loyal behavior (i.e., selected the same brand on three or more successive trials) by the end of the experiment.

Such findings may be explained by the fact that people tend to use cues systematically even when the cues possess no real predictive utility whatsoever (Dudycha & Naylor, 1966). In other words, "if a decision maker is placed in a situation where none of the cues available to him is of any value, he will *still* tend to pick out and use some of them as if they did have value [Blum & Naylor, 1968, p. 459]." This would apply not only to such trivial cue differences as that which exist between different letters of the alphabet, but also to such strongly differentiated cues as price and brand image. The implication here is that consumers may use cues to form impressions of product quality even when these cues may have little or no relationship to actual (i.e., objectively ascertainable) product quality. Some tentative support for this conjecture is provided by Green (1966), who found that a large proportion of consumers do not differentiate between relevant and irrelevant information.

A possible limitation of the current investigation is the type of *S*s used. It could be argued that the screening procedures insured the use of *S*s who were relatively "expert" consumers of the product. This would not only restrict the generality of our results, but may in part explain *S*s' tendency not to use price as a quality indicator. Both Scitovsky (1967) and Shapiro (1968) have hypothesized that consumers lacking experience with the product are more prone than expert consumers to use price in forming impressions of product quality.

Another possible limitation was the use of a fourth brand of beer in the composition-differences-absent cells. In retrospect, it might have been more appropriate to have randomly assigned each of the 68 *S*s in these four cells to one of the three experimental beers.

In conclusion, to the extent one can generalize from this experiment, product, and sample, the major findings of this investigation were as follows:

1. Price does exert an effect upon the perception of product quality when it is the only cue permitted to vary.
2. Price does not operate as an indicant of product quality when embedded in a multicue setting.
3. Brand image (as mediated by brand name) does affect perception of quality, especially for brands with strong positive images.
4. Given substantial differences in product composition, and on the basis of taste and aroma cues alone, beer drinkers are able to discriminate quality differences in the expected direction across different brands of beer.
5. Consumers may use cues systematically even when these cues possess little or no predictive power.

Clearly, attempts should be made to replicate these findings with different types of *S*s (e.g., nonexperts), incorporating different cues and treatment combinations (e.g., using several beers, each at different price levels, or pouring ultrapremium beer into cups bearing a regional inexpensive beer brand name) and using different types of products. Given the importance of quality perception in consumer decision processes,[3] such effort seems well justified.

---

[3] For example, results of a recent study indicated that the extent to which housewives perceive differences to exist across the various brands comprising a product class is strongly related to brand loyalty: the greater the perceived quality differences, the greater the likelihood of strong brand loyalty (Jacoby, 1970).

REFERENCES

ALLISON, R. I., & UHL, K. P. Influence of beer brand identification on taste perception. *Journal of Marketing Research*, 1964, 1, 36–39.

BLUM, M. L., & NAYLOR, J. C. *Industrial psychology*. New York: Harper and Row, 1968.

COX, D. F. The measurement of information value: A study in consumer decision-making. In William S. Decker (Ed.), *Emerging concepts in Marketing*. Chicago: American Marketing Association, 1962.

COX, D. F. (Ed.) *Risk taking and information handling in consumer behavior*. Cambridge: Harvard University Press, 1967.

DUDYCHA, L., & NAYLOR, J. C. Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance*, 1966, 1, 110–128.

ENIS, B. M., & STAFFORD, J. E. Consumer's perception of product quality as a function of various informational inputs. In Philip R. McDonald (Ed.), *Marketing involvement in society and the economy*. Chicago: American Marketing Association, 1969.

FROMAN, R. Marketing research. In J. H. Westing

(Ed.), *Readings in marketing.* Englewood Cliffs, New Jersey: Prentice Hall, 1953.

Garbor, A., & Granger, C. W. J. Price as an indicator of quality: Report on an inquiry. *Economica*, 1966, **33**, 43–70.

Green, P. E. Consumer use of information. In Joseph Newman (Ed.), *On knowing the consumer.* New York: Wiley, 1966.

Jacoby, J. An attitudinal model of multi-brand loyalty: Preliminary results and promotional strategies. Paper presented at the Spring Conference on Research Methodology, American Marketing Association, New York Chapter, New York, May 14, 1970.

Leavitt, H. J. A note on some experimental findings about the meaning of price. *Journal of Business*, 1964, **27**, 205–210.

McConnell, J. D. The price-quality relationship in an experimental setting. *Journal of Marketing Research*, 1968, **5**, 300–303. (a)

McConnell, J. D. The effect of pricing in an experimental setting. *Journal of Applied Psychology*, 1968, **52**, 331–334. (b)

Myers, A. H., & Reynolds, W. H. *Consumer behavior and marketing management.* Boston: Houghton Mifflin, 1967.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, **17**, 776–783.

Scitovsky, R. Some consequences of the habit of judging quality by price. *The Review of Economic Studies, 1944–45*, **12**, 32. Reprinted in Perry Bliss (Ed.), *Marketing and the behavioral sciences* (2nd ed.). Boston: Allyn & Bacon, 1967.

Shapiro, B. P. The psychology of pricing. *Harvard Business Review*, 1968, **41**, 14–16, 18, 20, 22, 24–25, 160.

Stafford, J. E., & Enis, B. M. The price quality relationship: An extension. *Journal of Marketing Research*, 1969, **6**, 456–458.

Tucker, W. T. The development of brand loyalty. *Journal of Marketing Research*, 1964, **1**, 32–35.

Tull, D. S., Boring, R. A., & Gonsior, M. H. A note on the relationship of price and imputed quality. *Journal of Business*, 1964, **37**, 186–191.

White, I. S. The perception of value in products. In Joseph Newman (Ed.), *On knowing the consumer.* New York: Wiley, 1966,

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Zajonc, R. B. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, 1968, **9**, (2, Pt. 2).

(Received August 27, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 580–585

EFFECT OF FITTINGNESS, TYPE OF GOODS, AND TYPE OF SLOGAN ON BRAND AWARENESS[1]

SASI MISRA[2] AND SUBHASH JAIN[3]

Indian Institute of Management, Ahmedabad, India

Seventy-two male *S*s were run in an experiment testing the hypotheses: (*a*) fitting brand names would produce greater brand awareness than nonfitting brand names; (*b*) brand awareness would be superior for shopping goods than for convenience goods; and (*c*) an unqualified slogan describing a particular product would show better brand awareness than a qualified slogan describing the same product, although the superiority of unqualified slogans with respect to brand recall would tend to dissipate over time. The results confirmed Hypotheses a and b, but failed to support Hypothesis c.

In India, brand names such as Aspro for headache pill and Dalda for hydrogenated vegetable oil are household words. Similarly, one might reasonably argue that Kleenex for paper towels and Levis for blue jeans are effective American vocabulary. It is a matter of common observation that some successful brand names tend to stand for the name of the product itself. Despite the fact that in a free market various other brands of the same product are sold, consumers are often not so aware of their existence. The near omnipresence of certain brands of consumer products has thus facilitated effective monopoly and has reduced the economic concept of a perfect competition to the status of a myth. Although the paramount importance of brand awareness on consumer behavior has been long realized, empirical studies of brand awareness per se are very recent.

Using three letter nonsense syllables of varying meaningfulness value as brand names, Kanungo and Dutta (1966) found that their Indian *S*s, male and female, showed greater memory for highly meaningful brand names than for brand names of low meaningfulness. In addition, they found superior brand awareness for high-utility products than for low-utility products. The product-utility variable, operationalized in terms of assumed need of a particular product for the individual, included three different categories of consumer products: products used by males, by females, and by both. Recent extensions (Kanungo, 1968, 1969) using Canadian *S*s reported fittingness or appropriateness of brand names as a better predictor of differences in learning of brand-product association. Cultural differences in brand recall were also noticed for the three different product-utility categories employed in these experiments.

One purpose of the present study is to test the generality of the effect of fittingness of brand names on brand awareness using Indian *S*s who not only belong to a different cultural milieu but also to multiple linguistic backgrounds. It is of some importance to note that most educated people in India speak English along with one or more regional languages. Due to centuries of the former British rule many facets of Indian commercial operations, including brand naming and advertising, continue to be in English even though India gained its independence from England more than 23 yr. ago.

The specific hypothesis at stake is whether brand awareness is better when brand names are fitting then when they are nonfitting. An adequate test of this hypothesis requires that potentially fitting brand names should be selected from *S*s' everyday associations of the generic products whose specific brand names are sought. A method for selecting brand names of varying degrees of fittingness in terms of the above operational criterion was proposed by Kanungo (1968).

---

[1] The authors gratefully acknowledge the comments and suggestions of R. Kanungo, McGill University, Canada, and Dale Berger, Claremont Colleges, California, in the preparation of the manuscript. Thanks are due to R. Balachandra, Indian Institute of Management, Ahmedabad, India, for making the students of his course available for this research and to Gargi Munshi and K. Kesavankutty for their help.

[2] Requests for reprints should be sent to Sasi B. Misra, Indian Institute of Management, Vastrapur, Ahmedabad, 15, India.

[3] Now at the University of Dayton, Dayton, Ohio.

A second purpose of the present study is to determine the comparative influence of two types of products, convenience goods and shopping goods, upon the strength of brand-product association. Convenience goods are "those consumers' goods which the customer usually purchases frequently, immediately, and with the minimum effort in comparison and buying. . . . These articles are usually of small unit value . . . [Alexander, 1960, p. 11]." Shopping goods are "those consumers' goods which the customer in the process of selection and purchase characteristically compares on such bases as suitability, quality, price, and style [Alexander, 1960, p. 21]." Thus, it may be reasonable to speculate that the advertisements of shopping goods would be relatively more salient or important than the advertisements of convenience goods for the buyer making his purchase decisions. While purchasing convenience goods, the buyer is presumably making relatively unimportant decisions involving little risk, whereas in purchasing shopping goods, he is making more important decisions involving greater risks and, hence, attempts to be aware of all possible alternatives. It is therefore expected that brand awareness would be superior for shopping goods than for convenience goods.

Finally, this article also examines the so-called delayed action effects of two types of slogans, unqualified versus qualified, on brand awareness. In the attitude change literature the former refers to one-sided communication and the latter to two-sided communication. Simply put, an unqualified persuasive communication contains arguments that only favor its conclusion, whereas a qualified communication adds some secondary qualifications and reservations against the main argument. Extending the postwar studies carried out by Hovland and his associates of the Yale Communication and Attitude Change Program (Hovland, Lumsdaine, & Sheffield, 1949; Hovland & Weiss, 1951) Papageorgis (1963) predicted that the persuasive impact of a qualified message would decay more slowly than that of an unqualified (one-sided) message. However, despite their differing decay rates the net impact may be the same due to greater initial impact of the unqualified message even though it declines faster than the qualified message. Papageorgis further reasoned that his predictions concerning opinion change would also apply to learning and retention of messages in the communication. This is because learning and recall of content are positively related to opinion change (McGuire, 1957). The results confirmed Papageorgis's prediction concerning retention of the contents of persuasive communication, but provided little support for his hypothesis regarding opinion change. In the context of advertisements through formal communication systems such as radio and television commercials, pamphlets, and newspaper advertisements, one encounters many instances of qualified as well as unqualified slogans characterizing various products. For example: "Things go better with coke" can be classified under unqualified advertisements, whereas the commercial for the dog food, Alpo, which claims to be "all meat" and, hence, costlier than others, can be classified under qualified advertisements. The purpose of the present study is to investigate the relative impact of these two types of slogans on brand awareness. Thus, it may be tentatively hypothesized that products described by unqualified slogans would show greater brand awareness immediately after their exposure as compared to products described by qualified slogans. However, the net impact of the two types of slogans may be the same after a delayed time interval of appreciable length.

In sum, the present experiment examines the effect of: (*a*) fitting versus nonfitting brand names, (*b*) shopping versus convenience goods, and (*c*) qualified versus unqualified slogans on immediate brand awareness and its persistence over time.

METHOD

Subjects

The *S*s were 80 adult males of the Indian Institute of Management at Ahmedabad, India. These *S*s came from different parts of India and were participating in 2-yr. educational program of business management. None had participated in any kind of psychological experiment before. Data for four *S*s had to be discarded because they failed to complete the experiment. In order to equalize condition *N*s, four more randomly selected *S*s were eliminated leaving 72 for final analysis.

Materials

Products. Two categories of consumer goods were chosen, convenience and shopping goods. Eight products from each of these categories commonly used by both males and females were selected.

Brand names. For each of the 16 products two brand names were prepared, fitting and nonfitting. The procedure used for the selection of brand names was much the same as that employed by Kanungo (1968). Briefly, 16 adult males and 16 adult females of the Indian Institute of Management at Ahmedabad, India, were asked to give within a 40-sec. period as many relevant associations as they could to each of the 16 products. For each product a separate sheet of paper was provided. The *S*s were instructed not to give any existing brand names that they might be aware of as their associative responses. They were also asked to give all responses in English. All responses for each product were found to be meaningful words.[4] The most frequent (occurring 15 times or more) and rarest (occurring only once) responses were then determined. From those response words, four brand names, two fitting and two nonfitting, for each product were constructed.[5] An additional 20 adults, 10 males and 10 females, were given a list of 64 brand names (for 16 products) in random order without any mention of the product they might represent. These *S*s were asked to rate each brand name on a 5-point scale with ends labeled "very meaningful" and "meaningless." The mean meaningfulness value (m′) for each brand name was calculated from these ratings following which one fitting and one nonfitting brand name of least different meaningfulness value for each product was selected. This procedure of brand name selection was followed in order to minimize the possible confounding of meaningfulness with the fittingness variable examined in the present study. As far as possible, fitting and nonfitting names were designed to resemble highly in sound, form, or meaning a frequent and a rare response, respectively.

Slogans. Slogans for each of the 16 products were prepared in qualified and unqualified versions. An unqualified slogan consisted of a short message favoring the use or purchase of a particular brand, the brand name being either fitting or nonfitting for the product it advertised. A qualified slogan not only consisted of an identical message contained in an unqualified slogan, but also some secondary mention of a qualification or reservation against the main message. However, any negative implications of qualifying the message for the product it represented were minimized consistent with one's general expectations that advertisements do not normally stress conditions under which advertising appeals might not apply. The products, the brand names, their meaningfulness values, and the slogans are presented in Table 1.

Design

For each of the 16 products, there were two different brand names: fitting (F) and nonfitting (NF). Each of these brand names was presented either with an unqualified (UQ) or with a qualified (Q) slogan. Thus, for each product, four different advertisements were prepared F-Q, F-UQ, NF-Q, and NF-UQ. Copies of each of these 64 advertisements were printed in the manner suggested by Kanungo (1968). However, the Kanungo series of studies presented a picture of the appropriate product along with each advertisement. The present study did not.

Finally, 80 booklets were compiled. Each booklet contained one advertisement of each of the 16 products of two categories: convenience goods and shopping goods. All advertisements in a given booklet comprised one of the four kinds: F-Q, F-UQ, NF-Q, NF-UQ. The advertisements of shopping goods preceded the advertisements of convenience goods half of the time and followed them for the other half in a counterbalanced order. Additionally, the sequence of the eight advertisements within each product category in each booklet was randomized to eliminate order effect in presentation. The name and address of a fictitious advertising agency appeared on the front page of each booklet.

Procedure

The experiment was conducted during the regular classroom session. Each *S* was given a separate booklet and was asked to rate each advertisement in the booklet on each of the three 7-point scales: pleasant-unpleasant, good-bad, and ambiguous-clear. The three bipolar scales were printed on the answer sheets and were labeled at the ends. The intermediate labels were numbered but not verbally defined. The order of presentation of the three scales was randomized for each answer sheet. For each scale the positions of polar adjectives were reversed half of the time. Finally, each *S* rated the appropriateness of each brand name on a 6-point appropriateness scale. The *S*s were told that the advertising agency intended to study certain "psychological aspects" of brand naming and advertising in order to serve its clients better. Hence, the advertiser wanted to determine consumer reaction to these advertisements and brand names.

The *S*s were given approximately 2 min. to evaluate each advertisement. Immediately following the rating of all 16 advertisements, the booklet and the answer

---

[4] Interestingly enough, some responses were given in American slang terms, such as, cute, spic-n-span, etc. This is understandable considering the popularity of hollywood movies, American magazines, and paperbacks among Indian college students in general and students of this Institute in particular. Moreover, the students here seem to be somewhat more familiar with the American language because the Harvard Business School serves as the educational collaborator of the Institute, and virtually all members of its faculty obtained their higher degrees in the United States.

[5] Response words like "greasy" for hair-oil and "nuisance" for alarm clock, although frequent, were not considered in the construction of brand names for those products. The authors viewed it rather unrealistic and awkward to label any consumer product in obviously negative terms.

TABLE 1

PRODUCTS, SOURCE WORDS, BRAND NAMES, THEIR MEANINGFULNESS, AND SLOGANS

Product name	Brand name	m'	Slogan
1. Soap			
F	Cleno (Clean and Cleanliness)	2.80	Use Cleno [Purry] for cleaner and softer skin. [It's worth paying a little more.]
NF	Purry (Pure)	2.75	
2. Tea			
F	Flavora (Flavor)	3.05	Take Flavora [Treat] for delightful armoa and richness. [Close lid to keep it that way.]
NF	Treat (Treating)	2.95	
3. Talcum powder			
F	Fragrant (Fragrance)	3.40	Use Fragrant [Economy] for extra freshness. [However, it doesn't guarantee a brighter complexion.]
NF	Economy (Economical)	3.55	
4. Biscuits			
F	Crispy (Crisp)	2.70	Buy Crispy [Golden] for delicious taste. [But watch your weight.]
NF	Golden (Golden)	3.50	
5. Fountain pen			
F	Smooth-flo (Smooth & flow)	3.40	Use Smooth-flo [Perma] for the best in writing [although it dries a little slower.]
NF	Perma (Permanent)	2.30	
6. Tooth paste			
F	Smile Wel (Smile)	3.20	With Smile-Wel [Tastee] remove bad breath. [Soft brush recommended.]
NF	Tastee (Taste)	3.40	
7. Headache pill			
F	Pain Kil (Pain Killer)	3.45	Take Pain-Kil [Hope] for fast relief from headache. [But for recurrent headaches consult your physician.]
NF	Hope (Hope)	3.20	
8. Hair oil			
F	Welgrom (Groom)	2.85	With Welgrom [Fixo] ensure healthy hair growth. [For best results apply on dry hair.]
NF	Fixo (Fix)	2.95	
9. Radio			
F	Musica (Music)	3.10	Buy Musica [Leisure] for clearer reception [even if it consumes a little extra electricity.]
NF	Leisure (Leisure)	3.15	
10. Fabrics			
F	Kolory (Color)	1.80	Buy Kolory [Thicrap] for the latest in textile technology. [But you need to iron it for a sharp look.]
NF	Thicrap (Thick)	1.70	
11. Bicycle			
F	Rider (Ride)	3.45	With Rider [Fun] it's the performance that counts. [You wouldn't regret paying up to 4% more.]
NF	Fun (Fun)	3.35	
12. Crockery			
F	Delikate (Delicate)	2.95	Get Delikate [Exclusive] and play an elegant host. [It's as fragile as rest of the China.]
NF	Exclusive (Exclusive)	3.15	
13. Blanket			
F	Cozee (Cozy)	2.60	Have Cozee [Feel] for a cozy winter [although wool and silk don't feel the same.]
NF	Feel (Feeling)	2.75	
14. Electric fan			
F	Kool (Cool)	3.50	Go for Kool [Smarty] in a hot summer. [It's worth investing just a few rupees more.]
NF	Smarty (Smart)	3.90	
15. Footwear			
F	Durable (Durable)	3.05	Use Durable [Shiney] and walk at ease. [Apply cream polish for a sharper look.]
NF	Shiney (Shine)	3.10	
16. Alarm clock			
F	Wek-up (Waking up)	2.95	With Wek-up [Signale] you will never miss an appointment [though it looks a little old-fashioned.]
NF	Signale (Signal)	3.10	

Note.—Items 1–8 are convenience goods, and Items 9–16 are shopping goods. The word(s) next to each brand name is the associative response word(s) from which the brand name was derived. Nonfitting brand names and additional information used for qualified slogan conditions are bracketed. Abbreviations: F = fitting, NF = nonfitting, m' = meaningfulness value.

sheets were collected, and S was given a recall sheet wherein all the 16 products were listed in alphabetical order. He was asked to recall as many brand names as possible and write them beside the name of the products with which they were associated. Two delayed recalls—after 24 hr. and after 7 days—of brand names were taken from the same Ss. The procedure followed during delayed recalls was similar to that followed during immediate recall.

RESULTS AND DISCUSSION

There were no significant differences among groups in their evaluative ratings of advertisements on the three semantic differential-type scales mentioned earlier. As a check on the fittingness manipulation, Ss rated each brand name on a 6-point appropriateness scale with ends labelled "1—very inappropriate" and "6—very appropriate." Results indicated that the F brand names were rated as more appropriate ($M = 3.96$) than the NF brand names ($M = 2.92$, $t = 2.03$, $p < .05$).[6] This suggests that the experimental manipulation of fittingness was effective.

The recall scores for individual Ss in this experiment were exceedingly high as compared to the low-recall scores reported by Kanungo (1968). No S showed complete absence of recall in any experimental condition.

Effects of Fittingness, Types of Slogans, and Delayed Action Effects

In order to determine the effects of fittingness, slogans, and time variables on brand awareness, a 2 × 2 × 3 analysis of variance was run on the recall of brand names. The three classifications were: fitting and nonfitting brand names; unqualified and qualified slogans; immediate recall, recall after one day, and recall after 7 days (see Table 2).

The main effect of fittingness was highly significant ($F = 31.69$, $p < .001$). The results undoubtedly support the earlier findings that a fitting brand name is retained better than a nonfitting brand name. No other main effect or interaction was significant. However, the brand recall for advertisements with unqualified slogans (overall $M = 12.53$) was slightly higher than those with qualified slogans (overall $M = 11.72$, $F = 3.26$, $p < .10$). For the immediate recall measure, brand awareness for advertisements with unqualified slogans was higher ($M = 12.64$) than those with qualified slogans ($M = 11.72$), but this difference was statistically insignificant, thus failing to confirm the authors' hypothesis that the immediate impact of unqualified slogans would be greater than that for qualified slogans.

TABLE 2

ANALYSIS OF VARIANCE OF RECALL SCORES

Source of variation	df	MS	F
Fittingness (F)	1	340.01	31.69**
Type of Slogan (Q)	1	35.05	3.26*
Time (T)	2	1.56	.15
F × Q	1	18.37	1.71
F × T	2	1.24	.09
Q × T	2	1.05	.12
F × Q × T	2	6.16	.57
Within cell (error)	204	10.73	
Total	215		

* $p < .10$.
** $p < .001$.

One of the striking findings in this experiment is that there was very slight forgetting of brand names as a function of time. It is possible that previous recall tests might have aided memory during the subsequent tests. A more adequate test of delayed action effect would have been to measure brand recall after 3 wk. or so. Unfortunately, for various reasons the same Ss were not available for such a test. Alternatively, one might use different Ss for each condition of the time variable or split brand names of products to be remembered into three subgroups and test for brand recall of one subgroup at a time.

Effect of Product Types

In another three-way design types of goods, fittingness, and the effect of time were considered as fixed factors. The analysis of variance may not be reported in detail, but only two systematic sources were significant: fittingness and types of goods, with Fs of 51.2 and 5.52, respectively. The latter F ($p < .05$) reflects greater weighted mean recall of brand names of shopping goods ($M = 6.75$) as compared to brand names of convenience goods ($M = 6.32$). This small effect may have less practical importance, but its statistical significance does testify to the high intraindividual reliability. It supports our hypothesis that brand awareness for shopping goods exceeds that for convenience goods.

---

[6] All p values are two-tailed.

CONCLUSION

The results of this experiment suggest that to create better brand awareness, advertisers of a new product should choose fitting brand names for consumer goods. Second, the failure to establish clearly the differential roles of unqualified and qualified slogans on brand awareness calls for further research in this area. There may be at least two reasons why these data did not bear out the predictions concerning the differential role of slogans. First, it may be argued that the qualified versus unqualified slogan effect is confounded with long versus short slogan effect. A second criticism suggests that the present research did not control for implicit versus explicit qualifications used in the qualified slogans. It may be difficult to eliminate these confounding variables, but some control for these kinds of confounding should be sought in future studies.

Finally, only males were used in this study, despite the fact that certain parameters were previously established on both males and females. This is simply because the experiment was carried out in an institution where there are very few female students. Although it would be imprudent to generalize our results for females, we are encouraged by the fact that our results paralleled some of the findings of earlier work using both males and females.

REFERENCES

ALEXANDER, R. S. *Marketing definitions.* Chicago: American Marketing Association, 1960.

HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication.* Princeton: Princeton University Press, 1949.

HOVLAND, C. I., & WEISS, W. The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 1951, **15**, 633–650.

KANUNGO, R. N. Brand awareness: Effects of fittingness, meaningfulness, and product utility. *Journal of Applied Psychology*, 1968, **52**, 290–295.

KANUNGO, R. N. Brand awareness: Differential roles of fittingness and meaningfulness of brand names. *Journal of Applied Psychology*, 1969, **53**, 140–146.

KANUNGO, R. N. & DUTTA S. Brand awareness as a function of its meaningfulness, sequential position, and product utility. *Journal of Applied Psychology*, 1966, **50**, 220–224.

MCGUIRE, W. J. Order of presentation as a factor in "conditioning" persuasiveness. In C. I. Hovland et al. (Eds.), *The order of presentation in persuasion.* New Haven: Yale University Press, 1957.

PAPAGEORGIS, D. Bartlett effect and the persistence of induced opinion change. *Journal of Abnormal and Social Psychology*, 1963, **67**, 61–67.

(Received July 31, 1970)

ournal of Applied Psychology
971, Vol. 55, No. 6, 586–588

CHARACTERISTICS OF NONRESPONDENTS AMONG WORKERS

MARTIN J. GANNON,[1] JOSEPH C. NOTHERN, AND STEPHEN J. CARROLL, JR.

Behavioral Science Division, Department of Business Administration, University of Maryland

Given the fact that surveys frequently have response rates below 50%, it is important to know the characteristics of nonrespondents. Comparisons of respondents and nonrespondents in survey research have indicated that females and those with higher educational levels have higher response rates. The present study confirmed these findings and also indicated that surveys of workers seem to be biased toward more stable, older, more effective employees. The *S*s were 552 checkers working in 14 stores of a supermarket chain.

It is often necessary to conduct surveys of employed persons. Surveys are common for determining worker attitudes about various job conditions, unionism, or a variety of social and political issues such as integration or the regulation of industry. Such surveys are sponsored or conducted by firms, government institutions, or behavioral scientists investigating some aspect of employee behavior. Four methods are in common use, namely, to conduct the survey on the job, through a visit to the worker's home, through a mail questionnaire completed at the place of employment, or through a mail questionnaire sent to the worker's home.

The first method, the completion of the survey on the job, is probably used most frequently because of the higher response rate from the population or sample of the population chosen for study. However, this approach can be so disruptive of work that it might be impossible or too time consuming for those conducting the survey. Also, it is possible that individuals will be less candid in the work situation than they would be in their homes. In addition, the costs of lost production while the survey is being conducted may be significant.

Surveys involving visits to workers' homes have been used primarily when attitudes of workers in general rather than those of workers in a particular organization have been the focus of the study. It is obvious from cost considerations why this is the case. In addition, research indicates that individuals have a tendency to give more socially acceptable answers in the presence of an interviewer than they do on a questionnaire (Hyman, Cobb, Fredman, Harts, & Stember, 1960). Mail surveys conducted at the work place also involve the cost of lost production to the organization. The mail survey to the worker's home has the advantage of not being disruptive of work activities and being less costly to the organization in terms of lost production but, similar to the work place mail survey, has the disadvantage of the nonresponses from a certain proportion of the group under study. (For further discussions, see Sudman, 1954 and Parten, 1950.)

In the situations where it is necessary or advantageous to conduct mail surveys of workers for any of the reasons previously cited, it is desirable to know how the nonrespondents will bias the returns. This information obviously will assist the researcher in formulating any conclusions based on the data. To contribute to this end, the authors have obtained some information on the characteristics of nonrespondents in a mail survey of workers in a supermarket chain and these data are presented in this report.

METHOD

Survey questionnaires were given to 552 checkers in 14 food stores of a supermarket chain in Washington, D. C., along with a stamped return envelope addressed to the authors. The questionnaire was designed to identify workers' attitudes about supervisory practices, the social climate, and personnel policies in the firm. Although the questionnaires were not to be signed, each was coded by means of a small ink mark. The firm was furnished only summary data and not any information on specific individuals. Research has demonstrated that using nonanonymous questionnaires does bias the

---

[1] Requests for reprints should be sent to Martin J. Gannon, Behavioral Science Division, Department of Business Administration, University of Maryland, College Park, Maryland 20742.

TABLE 1

PREDICTORS OF RESPONSE RATES AMONG WORKERS

Item	Respondent %	Nonrespondent %	N	χ^2
Education				
Under 11 yr.	57.4	42.6	61	
11 & 12 yr.	62.9	37.1	291	
13 & 14 yr.	64.6	35.4	116	
15 yr. & over	80.3	19.7	66	8.87*
Marital status				
Single	51.3	48.7	113	
Married	64.7	35.3	382	
Widowed, divorced, or separated	76.4	23.6	55	11.30**
Age				
Under 20 yr.	57.7	42.3	26	
20–24 yr.	66.9	33.1	133	
25–29 yr.	51.1	48.9	139	
30–49 yr.	69.4	30.6	232	
50 yr. and above	52.4	47.6	21	14.75**
Sex				
Male	56.8	43.2	192	
Female	66.0	34.0	352	4.56*
Supervisory ranking				
Twentieth percentile & under	53.4	46.6	73	
Twenty-first percentile & over	68.0	32.0	391	5.83*

Note.—N = 552; the N in each subtable does not total 552 due to incomplete information.
* $p \leq .05$.
** $p \leq .01$.

research results obtained (See Hyman et al., 1960; Metzner & Mann, 1952; and Yoder, 1962). Consequently, most surveys of workers rely on anonymous questionnaires.

The number of questionnaires returned was 348 of the 552 sent for a response rate of 63%. The 204 nonrespondents were compared to the respondents with respect to certain personal characteristics obtained from personnel records of the organization. Comparisons were made on the basis of years of education, length of service with the organization, marital status, sex, age, number of children, and supervisory rankings of effectiveness. The supervisory rankings were made by each store manager for the checkers in his store. These rankings were then converted to percentile scores so that a common effectiveness measure of the employees in the 14 stores could be used. Chi-square analysis was used to determine if any differences between the respondents and nonrespondents were statistically significant. In addition, each characteristic analyzed was compared to the response rate when every other characteristic was held constant. This procedure was necessary because all of the characteristics were related to each other.

RESULTS

Nonrespondents differed from respondents on five of the seven characteristics analyzed. Table 1 presents the results of the analysis. The table lists only those characteristics for which there was a statistically significant difference between the respondents and nonrespondents. Length of service and number of children were not found to be predictive of a nonrespondent and are not included in the table. Number of children initially had a relationship to the response rate, but this disappeared when marital status was held constant. In a similar manner, as indicated earlier, the five characteristics listed in Table 1 were found to be predictive of response rates among at least some categories of the other characteristics in the table when they were held constant.

As Table 1 indicates, a lower response rate is associated with lower levels of education, single status, male status, being younger or older than the middle-aged group (between 30 and 49 yr. of age), and ranking in the bottom fifth of all employees as measured by supervisory ratings. These findings confirm previous research (Ellis, Endo, & Armer, 1970; Goode & Hatt, 1952) that showed that females and those with higher educational levels have higher response rates.

In general the results of the analysis indicate that mail surveys of workers may be biased toward more stable, older, more effective employees. In addition, typical attitudes of female workers, to the extent that they differ from those of male workers, may be weighted too heavily by such surveys.

REFERENCES

ELLIS, R., ENDO, C., & ARMER, J. The use of potential nonrespondents for studying nonresponse bias. *Pacific Sociological Review*, 1970, **13**, 103–110.

GOODE, W., & HATT, P. *Methods in social research.* New York: McGraw-Hill, 1952.

HYMAN, H., COBB, W., FREDMAN, J., HART, C., & STEMBER, C. *Interviewing in social research.* Chicago: University of Chicago Press, 1960.

METZNER, H., & MANN, F. A limited comparison of two methods of data collection: The fixed alternative question and the open ended interview. *American Sociological Review*, 1952, **15**, 486–491.

PARTEN, M. *Surveys, pools, and samples.* New York: Harper, 1950.

SUDMAN, S. *Reducing the cost of surveys.* Chicago: University of Chicago Press, 1954.

YODER, D. *Personnel management and industrial relations.* Englewood Cliffs, N. J.: Prentice-Hall, 1962.

(Received September 3, 1970)

Journal of Applied Psychology
1971, Vol. 55, No. 6, 589–592

EFFECT OF PERSONAL SIGNATURE ON THE INITIAL RATE OF RETURN OF A MAILED QUESTIONNAIRE[1]

MARY B. KAWASH AND LAWRENCE M. ALEAMONI[2]

University of Illinois

An attempt was made to test the traditional rule of thumb that a personal signature on the cover letter is a *sine qua non* for yielding higher rates of return of mailed questionnaires. Three thousand and ninety-one members of the University of Illinois faculty were sent questionnaires concerning their uses of and attitudes toward audiovisual instructional materials. One group ($N = 1{,}546$) was sent a cover letter bearing the researcher's personal signature. A second group ($N = 1{,}545$) was sent a cover letter identical to the first group's except that the researcher's signature was a mimeographed facsimile. No difference was found between the initial return rates for the two groups even when the rank of the faculty member was identified. It was concluded that, for the University of Illinois faculty, a personal signature on the cover letter accompanying a questionnaire concerning audiovisual materials was no greater an inducement to respond than a cover letter bearing a mimeographed facsimile.

For decades the mailed questionnaire has been considered one of the most economical tools of researchers in a variety of fields. Like many favored procedures, however, it does have an Achilles' heel, nonresponse, that undermines the foundation stone of statistical inference—the assumption of random selection of respondents from the target population. Nonrespondents have been observed to differ from respondents on a number of characteristics (Hochstim & Athanasopoulos, 1970; Mayer & Pratt, 1966–1967; Schwirian & Blaine, 1966–1967; Scott, 1961; Wallace, 1954). Such a state of affairs can result in distorted estimates of population parameters (Schwirian & Blaine, 1966–1967; Shuttleworth, 1941; Wallace, 1954). Of no mean significance is the difficulty of ascertaining the extent to which these differential characteristics and those which have not been measured interact with the specific relationships being studied, thus affecting the researcher's acceptance or rejection of his hypotheses (Abeles, Iscoe, & Brown, 1954–1955; Robin, 1965; Schwirian & Blaine, 1966–1967; Suchman, 1962).

Procedures for securing replies from the nonrespondents have been proposed, some of which have been highly successful (Bachrack & Scoble, 1967; Eckland, 1965; Hochstim & Athanasopoulos, 1970; Robin, 1965; Sirken, Pifer, & Brown, 1960). These attempts, however, may be viewed as harassments and yield distorted information. It, therefore, indubitably behooves the researcher to construct his questionnaire in such a manner as to ensure the highest possible *initial* return rate.

A great deal of research has been undertaken throughout the past 30 yr. to determine the effects of a number of variables on return rate. Some of the variables studied are: (*a*) postage on return envelopes (Champion & Sear, 1969; Gullahorn & Gullahorn, 1963; Mayer, 1946; Robinson & Agisim, 1951); (*b*) form of mailing (Gullahorn & Gullahorn, 1963; Russell, Konrad, & Kaluzny, 1970); (*c*) length of questionnaire (Brown, 1965; Durant & Maas, 1956; Scott, 1961; Sletto, 1940); (*d*) appearance of questionnaire (Bender, 1957; Ford, 1968; Scott, 1961); (*e*) questionnaire rewards (Hancock, 1940; Kephart & Bressler, 1958; Robinson & Agisim, 1951); (*f*) guarantees of anonymity (Hoppe, 1952; Mason, Dressel, & Bain, 1961); and (*g*) characteristics of the cover letter, such as the sponsor (Filipello, Birg, & Webb, 1958; Roeher, 1963); appearance (Moore, 1941; Simon, 1967), and nature of appeal (Champion & Sear, 1969; Hammond, 1959; Linsky, 1965; Sletto, 1940). These are the variables that have been most frequently assessed systematically.

---

[1] The authors are indebted to Dennis Williams for modifying the cover letter for the questionnaire to obtain this data.

[2] Requests for reprints should be sent to Lawrence M. Aleamoni, Measurement and Research Division, University of Illinois, 307 Engineering Hall, Urbana, Illinois 61801.

It was with this latter category that the present study was concerned. Several authors (Frazier & Bird, 1958; Hammond, 1959; Longworth, 1953; Parten, 1950) have implied that the handwritten signature of the sender of the questionnaire is one of the accouterments requisite for obtaining a high rate of return. No one, however, at least as far as these authors have been able to ascertain, has restricted his research to the effect of this *single* cover letter variable on the *initial* rate of return. Linsky (1965), for example, designed a survey of members of a state nurses' association in which he systematically varied three characteristics of the cover letter, one of which was the personalization of the salutation and signature. Initial returns to a cover letter with handwritten salutation and signature were significantly greater than one containing a mimeographed "Dear Member" and mimeographed signature. On the other hand, a study by Emmett (see Scott, 1961) of TV viewers yielded no significant difference between initial return rates to letters that contained personal salutations and signatures and those addressed "Dear Viewer" and bearing a facsimile signature. Similarly, for Weilbacher and Walsh (1952), personalization of the greeting and signature yielded no increase in response rate over that for a "nonpersonalized" letter.

Only one study was discovered that varied signature only. Clausen and Ford (1947), dismayed at receiving a 23% return rate to a letter employing a "Dear Veteran" salutation and a facsimile signature of General Omar Bradley (celebrated World War II leader), randomly assigned their nonrespondents to one of each of the following conditions: impersonal salutation-personal signature, impersonal salutation-facsimile signature, personal salutation-facsimile signature, personal salutation-personal signature. Return rates for the second mailing did not differ significantly for the four groups. Clausen and Ford, however, maintain that "one cannot conclude that personalization of the initial letter would likewise have made no difference in response rate. The mere fact of follow-up is in itself a personalizing device [p. 501]."

TABLE 1

QUESTIONNAIRE RETURNS BY TYPE OF COVER LETTER SIGNATURE

Returns	Type of signature				z
	Personal		Mimeographed		
	N	%	*N*	%	
Total returns	441	28.52	420	27.18	.83
Completed	389	25.16	359	23.23	1.25
Uncompleted[a]	52	3.36	61	3.94	.92
Total mailed	1546		1545		

[a] Questionnaires in this group were those which were returned uncompleted but accompanied by an explanation as to why they could not be filled out. All questionnaires returned uncompleted but without an explanation were counted as nonresponses.

The present study was an attempt to clarify the nature of the effect of personalization, namely, that of the sender's signature, on the *initial* return rate.

METHOD

Printed questionnaires were mailed to all 3,091 faculty members of the University of Illinois who appeared on the University's mailing list. They were asked about their uses of, and feelings concerning, audiovisual instructional materials. The questionnaire contained 52 structured items with alternatives ranging from 2 to 10. Seven of the 52 items allowed the faculty member the additional option of recording his own alternatives. Fifteen hundred and forty-six of the cover letters accompanying the questionnaires were personally signed with the researcher's name by the researcher and an assistant over a 4-day period. This method was used in order to keep the signature from becoming sloppy. For 1,545 of the letters, the researcher's signature was a mimeographed facsimile. The letters' contents and salutations (Dear Faculty Member) were identical.

RESULTS

Table 1 presents the return rates for the two types of cover letter signatures. Slightly more than 28% of those faculty members who received the personalized signature letters returned the questionnaire, completed or uncompleted with an explanation. Of those members who were sent the facsimile signature letters, approximately 27% completed the questionnaire or returned it uncompleted with an explanation. Tests of significance for the difference between proportions were performed between the two groups for total returns, completed returns, and uncompleted returns.

TABLE 2

QUESTIONNAIRE RETURNS BY TYPE OF COVER LETTER SIGNATURE AND RANK OF THE FACULTY MEMBER

Returns by rank	Type of signature				z
	Personal		Mimeographed		
	N	%	N	%	
Professor					
Completed	132	33.93[a]	124	34.54	.18
Uncompleted	7	13.46	10	16.39	.43
Mailed	477	30.85	476	30.81	.02
Associate professor					
Completed	74	19.02	70	19.50	.17
Uncompleted	11	21.15	9	14.75	.89
Mailed	233	15.07	263	17.02	1.48
Assistant professor					
Completed	103	26.48	80	22.28	1.33
Uncompleted	7	13.46	7	11.48	.32
Mailed	304	19.66	269	17.41	1.61
Instructor					
Completed	38	9.77	27	7.52	1.08
Uncompleted	6	11.54	10	16.39	.74
Mailed	153	9.90	163	10.55	.59
Research professor					
Completed	1	.26	1	.28	.05
Uncompleted	1	1.92	0	.00	1.09
Mailed	7	.45	3	.19	1.44
Research associate					
Completed	3	.77	7	1.95	1.40
Uncompleted	4	7.69	2	3.28	3.29
Mailed	52	3.36	76	4.92	2.23
Graduate assistant					
Completed	0	.00	1	.28	1.08
Uncompleted	0	.00	1	1.64	.93
Mailed	6	.39	7	.45	.26
Other					
Completed	38	9.77	49	13.67	1.75
Uncompleted	16	30.80	22	36.06	.59
Mailed	314	20.31	288	18.64	1.16

[a] This percentage is based on the number of completed personal signature returns ($N = 389$).

No significant differences in initial return rates were found for any of these return categories.

Table 2 presents the mailed and return rates by the faculty member's rank. Tests of significance on the number of mailed letters at each rank indicates that a significantly larger proportion of research associates received the mimeographed signature. When the same tests were run on the return rates, however, only the uncompleted returns exhibited significance differences, and that was in favor of the personal signature.

DISCUSSION

Contrary to traditional rules of thumb, the presence of the handwritten signature of the researcher had little effect on the initial rate of return of a mailed questionnaire even when the rank of the faculty member was identified. Whether or not one may generalize to other populations of respondents and mailers and to other types of questionnaires are additional empirical questions.

The significant difference observed in the number of mailed questionnaires to research associates is not surprising as no attempt was made to randomly mail by rank. Also, the significant difference for the uncompleted returns is probably an artifact of the small number in that category.

One might speculate that the content of the letter was sufficiently personal to counteract the effect of the handwritten signature or that no *single* variable is very effective in inducing higher return rates. It may be that *combinations* of variables are more effective, such as the type of letterhead used and the prestige or authority of the signer. For example, if one wanted a higher rate of return from full professors, then the signature of a full professor may be needed. In any case, the small rates of return for both forms indicate that there are other variables, either singly or in combination, that are more powerful inducements for informants to reply.

REFERENCES

ABELES, N., ISCOE, I., & BROWN, W. F. Some factors influencing the random sampling of college students. *Public Opinion Quarterly*, 1954–1955, **18**, 419–423.

BACHRACK, S. D., & SCOBLE, H. M. Mail questionnaire efficiency: Controlled reduction of nonresponse. *Public Opinion Quarterly*, 1967, **31**, 265–271.

BENDER, D. H. Colored stationery in direct-mail advertising. *Journal of Applied Psychology*, 1957, **41**, 161–164.

BROWN, M. L. Use of a postcard query in mail surveys. *Public Opinion Quarterly*, 1965, **29**, 635–637.

CHAMPION, D. J., & SEAR, A. M. Questionnaire response rate: A methodological analysis. *Social Forces*, 1969, **47**, 335–339.

CLAUSEN, J. A., & FORD, R. N. Controlling bias in mail questionnaires. *Journal of the American Statistical Association*, 1947, **42**, 497–511.

DURANT, H., & MAAS, I. Who doesn't answer? *Bulletin of the British Psychological Society*, 1956, **29**, 33–34.

ECKLAND, B. K. Effects of prodding to increase mail-back returns. *Journal of Applied Psychology*, 1965, **49**, 165–169.

FILIPELLO, F., BIRG, H. W., & WEBB, A. D. A sampling method for household surveys. I. Panel recruitment for testing wines. *Food Technology, London*, 1958, **12**, 387–390.

FORD, N. M. Questionnaire appearance and response rates in mail surveys. *Journal of Advertising Research*, 1968, **8**, 43–45.

FRAZIER, G., & BIRD, K. Increasing the response of a mail questionnaire. *Journal of Marketing*, 1958, **23**, 186–187.

GULLAHORN, J. E., & GULLAHORN, J. T. An investigation of the effects of three factors on response to mail questionnaires. *Public Opinion Quarterly*, 1963, **27**, 294–296.

HAMMOND, E. C. Inhalation in relation to type and amount of smoking. *Journal of the American Statistical Association*, 1959, **54**, 35–51.

HANCOCK, J. An experimental study of four methods of measuring unit costs of obtaining attitudes toward retail stores. *Journal of Applied Psychology*, 1940, **24**, 213–230.

HOCHSTIM, J. R., & ATHANASOPOULOS, D. A. Personal follow-up in a mail survey: Its contribution and its cost. *Public Opinion Quarterly*, 1970, **34**, 69–81.

HOPPE, D. A. Certain factors found to improve mail survey returns. *Proceedings of the Iowa Academy of Science*, 1952, **59**, 374–376.

KEPHART, W. M., & BRESSLER, M. Increasing the responses to mail questionnaires: A research study. *Public Opinion Quarterly*, 1958, **22**, 123–132.

LINSKY, A. S. A factorial experiment in inducing responses to a mail questionnaire. *Sociology and Social Research*, 1965, **49**, 183–189.

LONGWORTH, D. S. Use of a mail questionnaire. *American Sociological Review*, 1953, **18**, 310–313.

MASON, W. S., DRESSEL, R. J., & BAIN, R. K. An experimental study of factors affecting responses to a mail survey of beginning teachers. *Public Opinion Quarterly*, 1961, **25**, 296–299.

MAYER, C. S., & PRATT, R. W. A note on nonresponse in a mail survey. *Public Opinion Quarterly*, 1966–1967, **30**, 637–646.

MAYER, E. N. Postage stamps do affect results of your mailing. *Printer's Ink*, 1946, **217**, 91.

MOORE, C. C. Increasing the returns from questionnaires. *Journal of Educational Research*, 1941, **35**, 138–141.

PARTEN, M. *Surveys, polls, and samples: Practical procedures*. New York: Harper, 1950.

ROBIN, S. S. A procedure for securing returns to mail questionnaires. *Sociology and Social Research*, 1965, **50**, 24–35.

ROBINSON, R. A., & AGISIM, P. Making mail surveys more reliable. *Journal of Marketing*, 1951, **15**, 415–424.

ROEHER, G. A. Effective technique in increasing response to mailed questionnaires. *Public Opinion Quarterly*, 1963, **27**, 299–302.

RUSSELL, S. D., KONRAD, R., KALUZNY, A. D. Influencing the respondent: An experiment in maximizing the response rate of mail questionnaires. *Sociological Abstracts*, 1970, **18** (3, Supplement 7), 22.

SCHWIRIAN, K. P., & BLAINE, H. R. Questionnaire-return bias in the study of blue-collar workers. *Public Opinion Quarterly*, 1966–1967, **30**, 656–663.

SCOTT, C. Research on mail surveys. *Journal of the Royal Statistical Society*, 1961, **124**, 143–205. (Series A)

SHUTTLEWORTH, F. K. Sampling errors involved in incomplete returns to mail questionnaires. *Journal of Applied Psychology*, 1941, **25**, 588–591.

SIMON, R. Responses to personal and form letters in mail surveys. *Journal of Advertising Research*, 1967, **7** (1), 28–30.

SIRKEN, M. G., PIFER, J. W., & BROWN, M. D. Survey procedures for supplementing mortality statistics. *American Journal of Public Health*, 1960, **50**, 1753–1764.

SLETTO, R. F. Pretesting of questionnaires. *American Sociological Review*, 1940, **5**, 193–200.

SUCHMAN, E. A. An analysis of "bias" in survey research. *Public Opinion Quarterly*, 1962, **26**, 102–111.

WALLACE, D. A case for -and against- mail questionnaires. *Public Opinion Quarterly*, 1954, **18**, 40–52.

WEILBACHER, W. M., & WALSH, H. R. Mail questionnaires and the personalized letter of transmittal. *Journal of Marketing*, 1952, **16**, 331–336.

(Received August 6, 1970)